Characterising temperate bacteriophages isolated from the microbiota of chronic respiratory disease.

Mohammad Adnan Tariq

PhD

2016

Characterising temperate bacteriophages isolated from the microbiota of chronic respiratory disease.

Mohammad Adnan Tariq
BSc (Hons) and MSc

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle for the degree of Doctor of Philosophy.

Research undertaken in the Faculty of Health and Life Sciences and in collaboration with Freeman hospital in Newcastle upon Tyne

September 2016

Abstract

Cystic Fibrosis (CF) is the most common autosomal recessive genetic disorder in the UK. A mutation in the CFTR gene, alters a Cl^- transporter protein resulting in dehydration at epithelial surface and a thick mucus layer that provides a nutrient rich environment ideal for opportunistic bacteria to colonise. Bronchiectasis (BR) is similar symptomatically to CF, with localised dilations and inflammation events of the bronchial tree that can be linked to lung trauma or allergy. *Pseudomonas aeruginosa* (Pa) is the most common opportunistic pathogen in CF that correlates to lowered lung function. *Burkholderia* cepacia complex (Bcc) species have been shown to be more problematic to clear than Pa, due to increased antimicrobial resistance and progression to Cepacia syndrome that can be fatal.

These opportunistic bacteria and their genome plasticity allows adaptation to the lung and can correlate to pathogenicity of chronic infection which is associated with poor clinical outcomes. As adaptation to the lung environment is such a key aspect of chronic bacterial infection in the lung this study focuses on the temperate phages infecting Pa and Bcc. Temperate phages once integrated into the bacterial chromosome have been shown in other bacterial backgrounds to aid bacterial adaptation through increasing rates of recombination. They have also been previously characterised to aid positive selection by carrying genes that aid bacterial survival, aiding evolution of both the bacterium and phage. The aim of this project was to characterise temperate bacteriophages chemically induced from 94 Pa (47 associated with CF and 47 with BR patients), 47 Bcc isolates (associated with CF patients).

This study focuses on 3 key areas. The first studies how phages induced from Pa and Bcc infect isolates from patients at different stages of chronic infection. All the bacterial isolates used in this study had at least one inducible lysogenic phage, in some instances polylysogens.

Infectivity through Pa phage cross infection showed that adult CF phage were more infective across the Pa panel. Paediatric CF phages showed an infection profile similar to that of Pa phage induced from BR patients >10 years after clinical diagnosis. Pa phage associated with BR patients <10 years after clinical diagnosis showed the least infection across the Pa panel.

Secondly, isolating and purifying temperate phages can be difficult and time consuming as phages that induce from the bacterial host in high numbers can mask secondary or tertiary phages. Importantly identifying individual phages can be complex, as a sensitive bacterial host is needed to isolate and propagate. This study uses novel genomic approaches to overcome this problem and separate mixed phage communities using k-mer abundance. Stratified phages have been annotated and compared for similarities that link to the clinical aetiology of the bacteria that carried that phage.

Finally, this study begins to map genetic traits that may aid phage longevity in a microbial system. A model has been previously characterised with lytic phages where Ig-like domains or BAM motifs may explain how phages bind to complex carbohydrates in mucus. When evaluating the incidence of BAM, Ig-like domains in temperate phage DNA sequences isolated in this study we identify an increase in BAM motifs, that may correlating to the evolution of disease in both CF and BR. This aids the proposed BAM model and the evolution of temperate phages interacting at mucosal surfaces.

Table of content

Table of Figure

Table of table

Acknowledgements

First and foremost, I would like to show gratitude to Dr Darren Lee Smith for giving me this opportunity. I would like to thank him for all of the support and reassurance he has given me with the practical work, and for the guidance he has given me during the provision of my thesis. I truly could not have asked for a better supervisor. I would like to thank him for all the extra effort he put in taking me to conferences and building me as a researcher. I would like to thank Darren for being extra tolerant and calm with me during times when I found it difficult and ultimately never giving up on me.

I would also like to thank Darren for giving me the opportunity to fulfil my lifelong passion, to embark on a scientific research and to hopefully become an academic.

I would like to thank Audrey Perry for providing me with the bacterial samples from the Freeman hospital. I am grateful to Dr Anthony De Soyza, Dr Stephen Bourke and Dr John Perry for supporting my research. I am appreciative to many of my friends and colleagues from A321 lab and D216 office both past and present. I would like to thank Paul Agnew for the support he has given me as a liaison to the graduate school. A distinct reference must go to Giles Holt and Francesca Everest who have been great lab partners and have become my good friends during my time at Northumbria University as a PhD student.

This thesis would not have been possible without the help and support from my family. I am indebted to my lovely father Tariq and humble mother Nasrin, the support they have given me throughout my PhD both financially and emotionally. I am grateful for my brothers Ahsan, Munnan and Masoom who have given me support throughout the tough times. I cannot thank enough my wife Nadia who has been the most tolerant to my rants and selfishness throughout my PhD. My sister-in-law Tasneem has been a lovely support to me and my lovely niece Ambreen and nephew Humza have kept me uplifted throughout my study. I owe this thesis to my son's Junaid Jawad and Haris, for whom I continued to struggle to see this thesis through. Finally, I have to acknowledge my faith which kept me strong and devote throughout my PhD, thank you all.

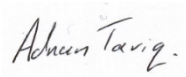
Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others. The work was in collaboration with Freeman Hospital in Newcastle upon Tyne.

Any ethical clearance for the research presented in this thesis has been approved. RE07-06-12940.

Name: Mohammad Adnan Tariq

Signature:



Date: 30-09-2016

List of Abbreviations

Abi: Abortive infection

ATP: Adenine triphosphate

AVOVA: One-way analysis of variance

Bcc: *Burkholderia cepacia complex*

BLAST: Basic Local Alignment Search Tool

bp: Base pair

BR: Bronchiectasis

$CaCl_2$: Calcium chloride

Cas: CRISPR-associated genes

cDNA: complementary DNA

CDS: Coding sequence

CF: Cystic Fibrosis

CFTR: Cystic Fibrosis transmembrane conductance regulator

cftr: Cystic Fibrosis transmembrane conductance regulator gene

contig: overlapping sequence

COPD: chronic obstructive pulmonary disease

CRISPR: Clustered regularly interspaced short palindromic repeats

CsCl: Caesium chloride

dH2O: distilled water

DNA: Deoxyribonucleic acid

ds: Double stranded

dsDNA: Double stranded DNA

EM: Electron microscopy

E. coli: *Escherichia coli*

EDTA: Ethylenediaminetetraacetic acid

g: Gram

hr: Hour

HHpred: Homology detection & structure prediction

HMM: Hidden Markov model

ICTV: International committee on taxonomy of viruses

IDBA-UD: Iterative De Bruijn Graph *de novo* Assembler - Uneven Sequencing Depth

Kbp: Kilo base pair

KCl: Potassium chloride

kDa: KiloDalton

k-mer: Short DNA sequence of fixed (K) length

LB broth: Luria bertani broth

LPS: Lipopolysaccharide

M: Molar

Mauve: Multiple genome alignment (program)

Mbp: Mega base pair

mg: Milligram

ml: Millilitre

mM: Millimolar

MOI: Multiplicity of infection

mRNA: Messenger RNA

NaCl: Sodium chloride

NaOH: Sodium hydroxide

NCBI: National Centre for Biotechnology Information

NFLX: Norfloxacin

NGS: Next generation sequencing

OD 600nm: Optical density 600 nanometres

ORF: Open reading frames

Pa: *Pseudomonas aeruginosa*

Pfam: Protein families

PFU: Plaque forming units

Phage: Bacteriophage

PHAST: Phage search tool

PLS-DA: Partial least squares regression - discriminant analysis

PriceTI: Paired-Read Iterative contig extension

RAST: Rapid annotation using subsystem technology

RFLP: Restriction fragment length polymorphism

RM: Restriction modification

RNA: Ribonucleic acid

RPM: Revolutions per minute

s: Seconds

SD: Standard deviation

SEM: Standard error of mean

SNP: Single Nucleotide Polymorphism

SPAdes: St. Petersburg genome assembler (program)

TA: Toxin antitoxin system

tRNA: Transfer RNA

UK: United Kingdom

UV: ultra violet

VelvetOptimiser: automated velvet *de novo* assembler (program)

w/v: Weight/volume

WHO: World Health Organisation

%: Percent

°C: Degree Celsius

µg: Microgram

µl: Microlitre

µm: Micrometer

Dedicated in loving memories of my Uncle

Abdul Rauf 1952 - 2013

1. Introduction

1.1 Cystic Fibrosis

CF is the most common autosomal recessive inherited disease in the UK, currently affecting over 10,000 people of whom only half will live past the age of 41 years. Currently there is no cure for people with CF where therapy aims to reduce the symptoms of the disease (NHS.uk, 2016). As a last resort, a CF patient may receive a lung transplant due to irreversible damage and lowered function. The option to transplant is recommended when the forced expiratory volume (FEV1) score is < 30 % of the predicted volume or if there is a rapid decline in lung function (Orens et al, 2006). Once the transplant has been successful, post-transplant care is vital for the maintenance of healthier lungs. Transplantation of the lungs means that the CF free lungs no longer produce thick mucus through limited dehydration. However, CF is still present in other vital organs such as the pancreas, that results in cystic fibrosis related diabetes, gastro intestinal tract, liver and gallbladder that causes highly acidic bile production and in males the testes (Cystic Fibrosis Foundation, 2014). In up to 15% of CF patients one of the first symptoms is blockage of the bowel due to viscous mucus preventing digestion or Intestinal atresia (O'Sullivan & Freedman, 2009). Other signs include greasy stool, abdominal bloating and poor weight gain that could also lead to malnutrition (O'Sullivan & Freedman, 2009). This increases the risk of the small intestines becoming infected by Gram negative bacteria which can also cause further malnutrition through competition for nutrients (Borowitz, 2005). Current research is mainly genetically oriented and focuses at the epithelial changes in CF patients as it is estimated that a 10% of normal CFTR function will cure the symptoms of CF-related disease (Ramalho et al, 2002).

CF is genetically linked to multiple mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. CF is termed a monogenic disease, as it is associated to modification of a single gene. CFTR is a 1480 amino acids protein that structurally displays six-membrane spanning glycoprotein domains (Sheppard & Welsh,

1999) with a nucleotide binding domain that is complimentary to interact with adenosine triphosphate (ATP) (Ames et al, 1990). The protein functions as a small conductance ATP and cyclic GMP-dependent chloride channel in the apical membranes of epithelial cells (Riordan et al, 1989). This protein assists the intracellular and extracellular movements of chloride ions. The difference in ions changes the salt and water content at the epithelial surface. The CFTR gene is 189 Kb in length and found at the q31.2 location in the long arm of chromosome 7, where expression is associated with the control of Cl^- ion channels that transport cellular Na^+ and Cl^- ions (Riordan et al, 1989). Polymorphisms in CFTR and alteration in the Cl^- ion channel subsequently affects the physiology of the lungs by stimulating the overproduction of thick mucus through reduced water content. Due to the different types of mutation in the CFTR gene, there are different severity and classes of mutation that correlate to the disease. These have been stratified into six CFTR mutation classes (MacDonald et al, 2007; Welsh & Smith, 1993). Currently 2009 CF mutations have been identified (Cystic Fibrosis mutation database, 2016), the most common mutation in Caucasians is the ΔF508, a 3 base pair deletion which leads to the deletion of the amino acid Phenylalanine. ΔF508 accounts for 66% of global mutations (Consortium, 1994) followed by G542X which is responsible for 2.4% of CF worldwide. The majority of mutations in *cftr* are due to missense (39.6%), frameshift (15.6%), sequence variation (13.4%), splicing (11.4%) or nonsense (8.3%) (Cystic Fibrosis mutation database, 2016).

Figure 1.1 illustrates the different classes of *cftr* mutation and how they alter the protein function. Class I defects result in ablation of CFTR at the atypical membrane (Welsh & Smith, 1993). Another example is the second most prevalent mutation G542X with a frequency of 2.4% (Zielenski & Tsui, 1995). Other mutations include W1282X and R553X and are regarded as nonsense mutations due to a premature stop codon.

Class II defects alter the processing of the CFTR protein, caused by proteasome degradation of the miss-folded protein found within the endoplasmic reticulum. This also results in the inhibition of *cftr* expression and localisation at the membrane (Welsh &

Smith, 1993). ΔF508 is an example of this mutation, mentioned earlier as the most prevalent, with a frequency of 66% (Zielenski & Tsui, 1995). Mutation N1303K is another example of a mutation that results in this class defect.

Class III mutations give rise to limited CFTR expression at the atypical membrane. This defect results in a gating defect, as CFTR cannot be activated due to missense mutations (Welsh & Smith, 1993). G551D is an example of a mutation with a population frequency of 1.6% (Zielenski & Tsui, 1995). G551S and G1349D are also examples of this type of mutational defect.

Class IV mutations result in limited expression of the activated CFTR with a decline in conductance as it hampers Cl^- ion movement (Welsh & Smith, 1993). These missense mutations result in the substitution of amino acids in the channel of the pore. The R334W mutation installs a change at codon 334 in the CFTR gene and results in a change in arginine to tryptophan (Gasparini et al, 1991). R117H is another mutation like R334W that alters the membrane spanning domain with an incidence frequency of 0.3% in the UK (Rosenstein et al, 1998). R117H mutation is an example of alternate splicing which occurs due to a poly-T tract variant at exon 9. Having 9T, 7T or 5T's skips exon 9 and can increase chance of presenting with CF (Chu et al, 1993). R347P are also examples of class IV based mutations.

Class V mutations result in the expression of native CFTR proteins at the atypical membrane however, in extremely reduced numbers compared to healthy individuals. An example of this type of mutation is 2789 + 5G->A (accounting for 0.1% of CF mutations worldwide (Dugueperoux & De Braekeleer, 2005)) and A455E.

Class VI is a result of increased turnover of CFTR protein at the atypical membrane however; it is unstable and is quickly degraded. An example of this type of CFTR classification is rescued ΔF508 and 120Δ23 (MacDonald et al, 2007). Other examples include N287Y, 4326ΔTC and 4279insA.

The severity of physiological symptoms associated with CF, decrease with the movement from class I to VI. Each class is linked to the degree of impairment assigned by each mutation to the CFTR protein. The class mutations have an effect on the

viscosity of the mucus, which in turn reflects the degree of mucus clearance. In one study CF patients with two class I mutations had a greater decline in lung function assessed by FEV1 and FCV scores, when compared to one class II mutation (Geborek & Hjelte, 2011). The genotype can therefore affect the phenotype within CF patients (Bonadia et al, 2014; Ferec & Cutting, 2012). However, in a clinical prospective, genotype does not accurately forecast the patients' long term outcome (Castellani et al, 2008).

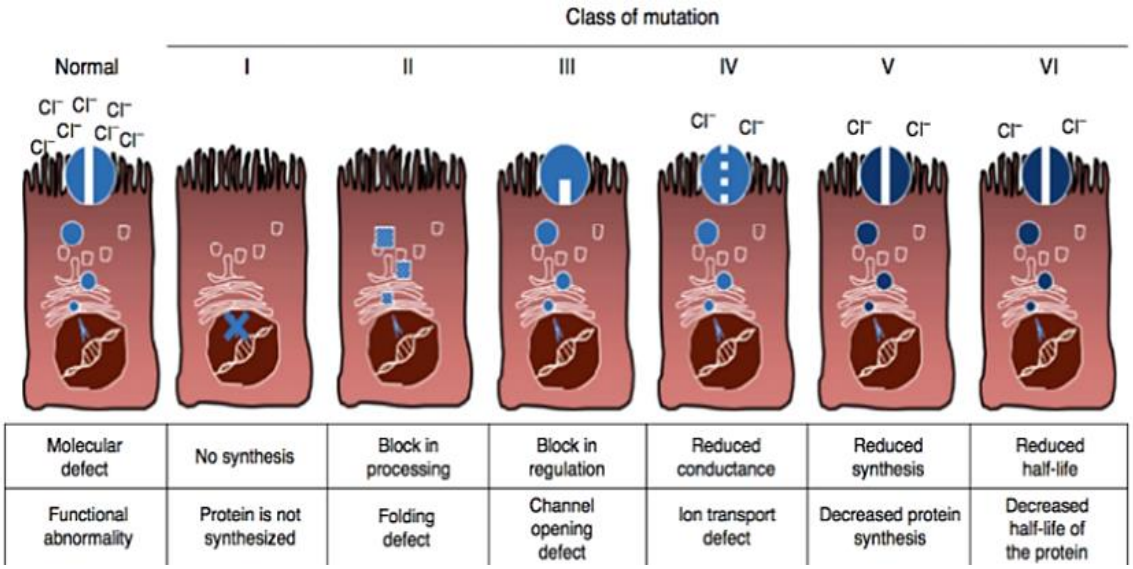


Figure 1.1 Cytic Fibrosis mutation classifications. The figure displays the location and effect of each class mutation. The image was taken from Quintana-Gallego *et al,* and altered (Quintana-Gallego et al., 2014).

Figure 1.1 shows a schematic diagram of the epithelial cell in the lung and illustrates where the class I – VI mutations effect the production of viable CFTR proteins. Classifying the mutation in this manner reveals that the three most prevalent mutations G542X, ΔF508 and G551D fall within Class I, II and III respectively. The type of mutation has an effect on the clinical outcome. A patient with a class V CFTR defect is less likely to have chronic bacterial infection of the lungs and is thus likely to have a higher life expectancy compared to someone with class I CFTR defect (Dugueperoux & De Braekeleer, 2005).

The altered ionic potential and pH can offer a nutrient rich environmental niche in which opportunistic bacterial pathogens can colonise where complex bacterial communities have been identified in the sputum of CF patients (Rogers et al, 2003). Impaired CFTR protein results in high Cl^- ions inside the epithelium and thus drives intake of H_2O, dehydrating the mucus. Repeated bacterial infection or continuous colonisation within the lungs results in epithelial inflammation and downstream scarring that inhibits the mucocilliary action of that area of the lung (Livraghi & Randell, 2007). Epithelial alterations through scarring stimulate further presence of mucus through lack of clearance increasing the risk of microbial colonisation and bacterial proliferation, which in turn causes further inflammation. These continuing rounds and exacerbation of symptoms increase scarring and diminished lung function. This is the main cause of the shortened life span of CF and Chronic respiratory disease patients (Callaghan & McClean, 2012; Gibson et al, 2003).

Pseudomonas aeruginosa, Burkholderia cepacia complexes*, Haemophilus influenzae* and *Staphylococcus aureus* are the main pathogen causing infections in lungs of CF patients (Mahenthiralingam et al, 2005; Valenza et al, 2008). It is difficult to clinically intervene and clear bacterial infections in the lung of CF patients (Hart & Winstanley, 2002) compared to Bronchiectasis patients, possibly due to a lifetime of re-occurring inflammation events which leads to damaged tissue, furthering poor drug delivery and pharmacokinetic profiles of prescribed antimicrobials (Athanazio et al, 2010). Some studies have shown that during periods of exacerbation there is a core microbiota in the lungs that is well conserved (Hauser et al, 2011; Rudkjobing et al, 2011). This therefore poses the question, as to why these patients have exacerbation periods, where one possibility is the involvement of viruses or the role of temperate bacteriophages that have the ability to alter the function of the cell.

Antibiotics are currently used to fight bacterial infections where subspecies of *Pseudomonas aeruginosa, H. Influenza* and *Burkholderia cepacia* species have developed antibiotic resistance due to mutation, adaptation, and genetic addition, thus

promoting evasion of treatment strategies. Genetic addition can occur through transposable genetic elements including; plasmids, bacteriophage and transposons. This study will look at the subversion of bacteria through the genomic addition via integration of temperate bacteriophages in *Pseudomonas aeruginosa* and the *Burkholderia cepacia* complex.

1.2 Bronchiectasis and its prevalence

CF is associated with the development of symptoms linked to bronchiectasis (BR) (Pasteur et al, 2000). While BR can be defined by the localised dilations and inflammation events of the bronchial tree that can be linked to lung trauma, in some instances allergic response and subsequent bacterial infection (Athanazio, 2012). The incidence rate of bronchiectasis in the UK is 1 in 1000 adults and 1 in 10,000 children. Over 12,000 patients were admitted to hospital in England from 2013 - 2014 due to bronchiectasis related symptoms and the majority of these patients were over 60 years of age (NHS.uk, 2015). BR is a long-term condition, which ultimately leads to excess mucus production.

1.2.1 Bronchiectasis and contributing factors

The onset of BR is multifactorial. Some are usually evident while other causes remain unknown; hence termed idiopathic BR. BR can occur if a patient has suffered from a viral or bacterial infection such as measles or whooping cough, lung cancer or the inhalation of toxic gases (Boyton & Altmann, 2016). Defects in mucocillary action (Morillas et al, 2007) and specifically autoimmunity diseases such as lupus erythematous (Higenbottam et al, 1980) can also bring about the onset of BR through the body attacking its own epithelia.

1.2.2 Polymicrobial nature of Bronchiectasis

The lungs of BR patients have been shown to be colonised with a polymicrobial community, illustrated by numerous metagenomics and 16S rRNA gene amplicon

sequencing studies (Delhaes et al, 2012; Erb-Downward et al, 2011; Fodor et al, 2012; Guss et al, 2011; Pragman et al, 2012; Willner et al, 2009). These infections usually include but are not limited to *Pseudomonas aeruginosa, Haemophilus influenza* and *Streptococcus pneumoniae*. Opportunistic bacteria take advantage of the environmental conditions in the lungs of these patients and subsequent to colonisation, can grow in a vegetative state. The bacteria can differentiate to grow as a biofilm making treatment more difficult (Ohgaki, 1994). The treatment strategies and intensity of antimicrobial therapy for clearing infection in BR patients depends on the severity and stage of the disease, but as in CF the option is to treat hard and treat early to limit chronic colonisation (Daniels, 2010).

1.2.3 Bronchiectasis relationship to Cystic Fibrosis

CF as mentioned in section 1.2 is symptomatically related to BR. The microbial nature of the lung between the two has been shown to be poly-microbial in a number of studies while there being a dominant species present (Cox et al, 2010; Purcell et al, 2014b; Rogers et al, 2004; Rogers et al, 2003; Sibley et al, 2006). A study comparing culture and sequence technique by amplification of hyper-variable region of 16s rRNA gene confirmed that the lung of both CF and BR are evidently more poly-microbial than previously assumed (Duff et al, 2013). However, it was seen that the lungs of BR patients had more operational taxonomic units (OTU) than CF patients (Duff et al, 2013).

1.3 Chronic obstructive pulmonary disease

Chronic obstructive pulmonary diseases (COPD) are those that give rise to chronic bronchitis and emphysema (NIH, 2016). COPD is predominantly the result of long term tobacco smoking and air pollution. This primarily damages, the lung through rounds of inflammation producing thick mucus that is difficult to clear by cilia action alone making it an ideal niche for bacteria to colonise and propagate (Ramos et al, 2014;

Rogers et al, 2013; Thornton et al, 2008; Voynow & Rubin, 2009). The lungs of COPD individuals are polymicrobial as shown by a number of metagenomics 16S studies (Delhaes et al, 2012; Erb-Downward et al, 2011; Fodor et al, 2012; Guss et al, 2011; Pragman et al, 2012; Willner et al, 2009). Including but not limited to *Pseudomonas aeruginosa, Burkholderia cepacia, Haemophilus influenza and Streptococcus pneumonia* have been found to be present in lungs of patients with COPD. Opportunistic bacteria take advantage of the conditions in the lungs of COPD patients and differentiate as biofilms that offer increased resistance to antibiotics and increase the difficulty of clearance by the host immune system alone, specifically in *Pseudomonas aeruginosa* (Hoiby et al, 2010).

1.4 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa (Pa) was first discovered in Paris by Carle Gessard in 1882. He noted that it fluoresced under UV light and was later described as an opportunistic pathogen in humans and plants (Gessard, 1984). He also noted that bandages of his patients would turn to green/blue and termed the pus as “blue pus” (Gessard, 1984). Pa is a Gram negative, coccobacillus bacterium; it is rod shaped and is commonly found in soil. Pa is known for its light green pyocyanine pigment that it secretes (Wilson et al, 1987). When Pa is grown in a culture vial and shaken vigorously the culture turns from a fluorescent yellow colour to green as it produces pyocyanin. Pyocyanin, a redox active phenazine based compound is part of the bacteria’s quorum sensing (QS) (Dietrich et al, 2006). Pyocyanin also plays a role in reducing iron from transferrin in low iron conditions, required for Pa growth (Cox, 1986). Pyocyanin is actually a blue pigment and Pyoverdine yellow and the combination of these two give the culture this green colour (Reyes et al, 1981). Some species of Pa can also produce pyomelanin and poyorubin, pale brown and pale red in colour respectively (Meyer, 2000).

In the chronic lung, Pa is the most common pathogen found (Bittar et al, 2008) and can prove to be problematic as it is difficult to clear by antibiotics (Sorde et al, 2011).

This is due to multi drug resistant strains of Pa and their ability to colonise and survive through limited metabolism as a biofilm. The genome size of Pa strain PAO1, first sequenced in the year 2000 was shown to be a 6.3 Mbp genome (Stover et al, 2000). Across strains the genome sequence of Pa varies between 5.2 - 7 Mbp (Klockgether et al, 2004) and this diversity in genome size is probably related to the receptiveness to transposable genetic elements.

Bacteriophages infecting Pa can have structure associated with filamentous, polyhedral or pleomorphic morphology (Ceyssens & Lavigne, 2010). Pf1 is an example of a filamentous phage of PAO1; it has a 15.7 Kbp genome. PAO1 also harbours a 35 kbp phage called MD8 and belongs to the *Siphoviridae* family. Phages infecting Pa predominantly belong to the order of *Caudovirales* comprising of three major families *Myoviridae, Podoviridae* and *Siphoviridae*. Phage phi CTX is an example of a *Myoviridae* phage. It encode a ctx gene in the Pa strain PA158 and has a 35.5 Kbp genome (Hayashi et al, 1990). Pa phage B3 and D3112 are examples of two *Siphoviridae* phages that infect the cell via adsorption to the IV pili (Roncero et al, 1990). These two phages are capable of genetic transposition and during excision package part of the host chromosome, aiding horizontal gene transfer (Morgan et al, 2002). F116 and H66 are part of the *Podoviridae* family, H66 has a genome size of 65.2 Kbp and F116 has genome size of 65.1 Kbp (Lammens & Lavigne, 2013; Maya et al, 2012). F116 is a transducing phage that adsorbs to the cell via the pili (Byrne & Kropinski, 2005; Pemberton, 1973).

A Pa strain resistant to all β lactam antibiotics was identified in Liverpool and termed the Liverpool epidemic strain (LES). LESB58 was found to contain five prophage regions and hypothesised to offer LESB58 a selective advantage (James et al, 2012; Winstanley et al, 2009). Phage integration correlates to an increased genome size of LESB58 which is 6.6 Mbp when compared to PAO1 6.3 Mbp. Gene addition and conversion of the bacterial cell through phage integration may be one of the reasons why LES is so pathogenic in a CF population. The LESB58 strain's temperate phages were shown to have individual biology and phenotypes for example different induction rates

under stimulus with norfloxacin and differential ability to induce spontaneously post infection (James et al, 2012). The lytic activities of the LESB58 temperate phages have previously been assessed *in vivo* over 2 years, and LESϕ2 and LESϕ4 were shown to have the most abundant free phage (James et al, 2015). The study also highlighted that the free phage densities exceeded LESB58 by 10-100 fold (James et al, 2015). This led the authors to infer that the lytic activity of the phages helped regulate the host density in the lung of CF patients (James et al, 2015). When the phages from the LESB58B were induced using norfloxacin and used to re-infect the susceptible host Pa (PAO1), it was identified that infection by LESϕ2 generated immunity against LESϕ3 and reduced susceptibility of LESϕ4, even though they were induced from the same strain (Winstanley et al, 2009). The same study also showed that the phages of LESB58 had a higher rate of spontaneous induction inferring their lack of stability in PAO1 when compared to LESB58. F. Jacob and J. Monod (1961) hypothesised early in the 1960's that a repressor molecule encoded in *E. coli* phage λ can turn off other infecting phage early gene expression and block integration and proliferation. This thus conforms to the model suggesting repressor molecules can prevent superinfection by other phage with similar genomes.

1.5 *Burkholderia cepacia* complex (Bcc)

Burkholderia is currently composed of over seventy species, most of which are associated with plants (Angus et al, 2014). The genus of *Burkholderia* was previously part of the *Pseudomonas* genus up until 1992 (Yabuuchi et al, 1992) when 7 species were moved into the *Burkholderia* genus. To complicate matters further multiple strains were transferred from the *Burkholderia* genus to the *Ralstonia* genus (Yabuuchi et al, 1995) illustrating the complexity and origin of this genus. Both 16S rDNA analysis (Reis et al, 2004) and multilocus sequence analysis on *acdS* (Onofre-Lemus et al, 2009) *gyrB*, (Tabacchioni et al, 2008) *rpoB*, (Tayeb et al, 2008) and *recA* (Payne et al, 2005) are a few examples of genes used to support the observations that the *Burkholderia* genus had two distinct groups of species. A study combining 16S rDNA and *atpD*, *gltB*, *lepA*

and *recA* genes to determine phylogenetic lineages on their *Burkholderia* species suggested that there might be a third main lineage (Estrada-de los Santos et al, 2013). This further implies that the *Burkholderia* genus is still evolving and is more than likely to be subject to future taxonomic changes.

The first lineage comprises: soil, plant and water associated species that tend not to be pathogenic. Members of this lineage have been illustrated to have symbiotic relationships with plants (Talbi et al, 2013) and bioremediation potential in degrading xenobiotics (Andreolli et al, 2011). The *Burkholderia* spp. have been shown to be able to use 2, 4, 5–trichlorophenoxyacetic acid (used in pesticides and herbicides) for its metabolism (Hubner et al, 1998). *Burkholderia* species have been known to aid plant growth hence they have been used to help crop yield and viability. *B. vietnamiensis* species have also been used to aid rice paddy fields in Asia (O'Sullivan et al, 2007). The second sub-lineage comprises: opportunistic bacterial pathogens, mammalian pathogens, plant pathogens and environmental species (Coenye & Vandamme, 2003).

This study focuses on specific Bcc species with the propensity to infect the lungs of CF patients. Bcc are Gram negative bacillus commonly found in water and soil. *Burkholderia* was first described in the 1950s as a pathogen of bacterial rot in onions by Burkholder (Burkholder, 1950). Bcc currently comprises of 20 sub species of which 10 species are commonly found in the lungs of CF patients hence termed *Bc complex* (Coenye et al, 2001; Cystic Fibrosis Foundation, 2014). Table 1.1 shows the species with their assigned genomovars (Mahenthiralingam et al, 2008).

Genomovar	Species
I	*Burkholderia cepacia*
II	*Burkholderia multivorans*
III A-D	*Burkholderia cenocepacia*
IV	*Burkholderia stabilis*
V	*Burkholderia vietnamiensis*
VI	*Burkholderia dolosa*
VII	*Burkholderia ambifaria*
VIII	*Burkholderia anthina*
IX	*Burkholderia pyrrocinia*
X	*Burkholderia ubonensis*

Table 1.1: *Burkholderia cepacia complex*- genomovar grades. The table lists the current members of the sub-species of the Bcc.

In CF patients 90% of the Bcc isolated from the lungs are *B. cenocepacia* or *B. multivorans* species (Mahenthiralingam et al, 2002). However, *B. multivorans* is becoming the most common species found of Bcc in the CF lung (LiPuma, 2010). The shift from *B. cenocepacia* to *B. multivorans* is possibly due to improved clinical management (De Boeck et al, 2004) and that the *B. multivorans* species isolated are largely unrelated suggesting environmental as opposed to patient to patient transmission (Turton et al, 2003). *B. multivorans* have been shown to undergo adaptation and evolution in the lungs of CF patients, such that mucoid to non-mucoid morphology changes have been described (Zlosnik et al, 2011; Zlosnik et al, 2008). Bcc has been shown to be more problematic to clear than *Pseudomonas aeruginosa*, due to its increased antimicrobial resistance. Although the lung function decline is only slightly greater with Bcc infections, infection with *B. dolosa* has been linked to increased lung decline and failure (Kalish et al, 2006).

The genome of *Burkholderia* species typically comprises circular chromosomes made up of multireplicons. *Burkholderia* species often contain 2-4 replicons which can vary in size; the largest replicon belongs to N2P5 with a size of 9.2 Mb (Viallard et al, 1998). *Burkholderia* species can also contain plasmids, as *B. cepacia* strain ATCC 25416 (genomovar I) has been characterised containing 3 chromosomes and a large plasmid, as the plasmid has four operons (Rodley et al, 1995), this also suggests high genomic plasticity.

1.6 Introduction to Bacteriophages

Bacteriophages are viruses that can infect and subvert their bacterial host. Identified originally by Frederick Twort in 1915 and later confirmed by Felix d'Herrelle in 1917 (d'Herelle, 1917; Twort, 1915). They were defined as filterable transmissible agents that can cause bacterial lysis. However, a recent study showed that by definition phage were filterable and because of this definition none filterable viruses have been missed and requires revision of the definition which classifies viruses as a whole. Mimivirus (although it is not a bacteriophage) is such an example of these giant viruses that are not filterable through a 0.2 µm filter, as it is around 0.8 µm in size (Raoult & Forterre, 2008). Phage sizes were first estimated using the filtration method and were estimated as being between 25 to 100 nm in size.

Using phage morphology to order and categorise has been the gold standard, these include long, short, contractile tails, non-contractile tails and whether they have an icosahedral capsid have all been visual characteristics used to classify phages. The most common way to observe phages for their morphological characteristic is via electron microscopy. However, taxonomically, viruses are categorised based on their nucleic acid content and morphology along with bacterial host range (ICTV, 2016). Viruses infect; plants, invertebrates, vertebrates, bacteria, algae, fungi, yeast and protozoa that form every clade of the tree of life. Figure 1.2a illustrates the different morphologies of viruses that infect bacteria. The guidelines of taxonomically naming phages aims to standardise

and remove ambiguities associated with names of phages. Recently the guidelines published recommended to remove phi, hyphens and “like” in the nomenclature of phages, as they add no informational value (Krupovic et al, 2016).

In *Pseudomonas aeruginosa* isolated from CF sputum, tail phages belonging to the family of *Myoviridae, Siphoviridae* and *Podoviridae* have been identified, notably temperate phages characterised thus far have tails (Ojeniyi et al, 1991) all of these families belong to the order of *Caudovirales*. Figure 1.2b shows an electron microscopy image of these families.

1.2a

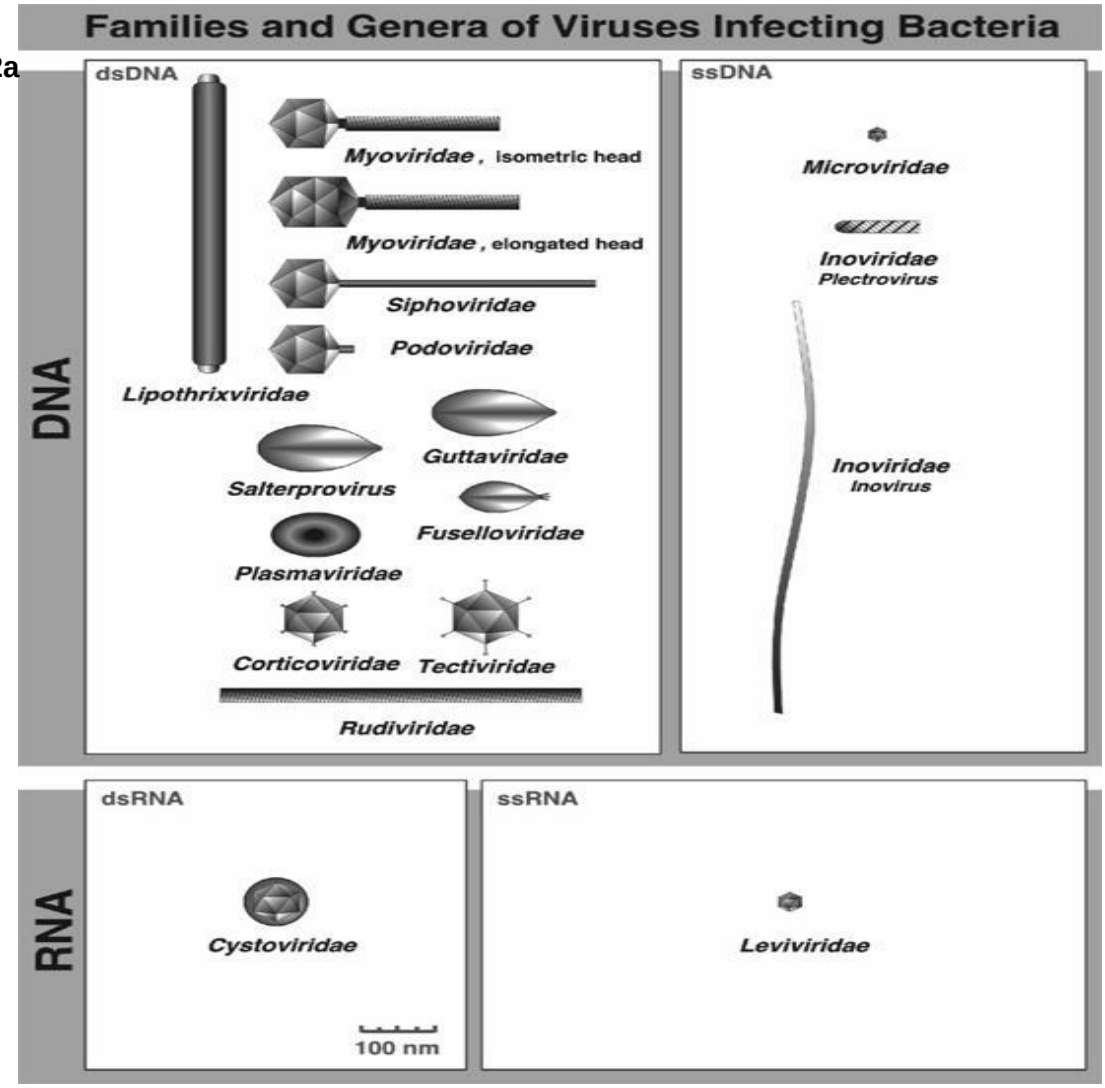


Figure 1.2a Viruses that infect bacteria. This diagram illustrates the type of viruses that can infect bacteria, image taken from Fauquet and Schrock (Fauquet C.M. & Schrock J.R., 2006).

1.2b

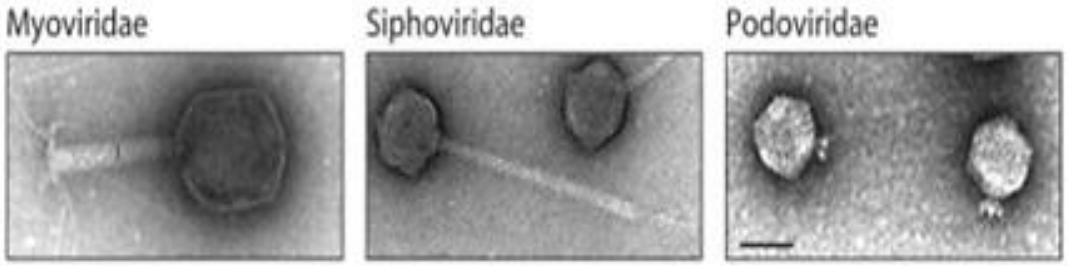


Figure 1.2b electron microscopy image of bacteriophage. The *Myoviridae* image if of a T4-like bacteriophage from Damien Maura and Laurent Debarbieux, institute Pasteur. The *Siphoviridae* is a HK97 phage and *Podoviridae* is a P22 phage from Laurence *et al,* (Lawrence et al., 2002). The black bar represents 50nm and is to scale for all the images. The images were taken from Krupovic *et al*, 2011 (Krupovic et al., 2011).

1.6.1 The life cycle of temperate bacteriophages

Escherichia coli infecting dsDNA phage Lambda or ʎ was the first phage to be characterised in detail, and has become the archetypal model for research on temperate phages. Tailed bacteriophages can be categorised as either temperate or lytic. Temperate phages can enter either of two life cycles on infection of a bacterial host. They can enter either a lysogenic pathway where the phage gDNA site integrates their genome into the chromosome of their bacterial host, mediated by integrase. Or they can also enter into the lytic pathway where they propagate and lyse the host cell to enter the environment. Upon infection the genome of the temperate phage is able to site specifically integrate into the bacterial chromosome. Lytic phages only infect to propagate and do not integrate their genomes into the host cell. The bacteriophage life cycle is controlled by gene expression and is discussed in chapter 1.6.5. Figure 1.3 illustrates the lytic and lysogenic life cycle. Viruses that only exhibit the lytic life cycle are sometimes termed virulent viruses as the inevitable outcome of a successfully infected host cell results in cell death. The following sections will isolate each stage of the key stages of temperate phage infection and describe in more detail: adsorption and therefore focus on the phage tail and virus structure (chapter 1.6.3), gene expression on infection and life cycle decision (chapter 1.6.5) through to lysis.

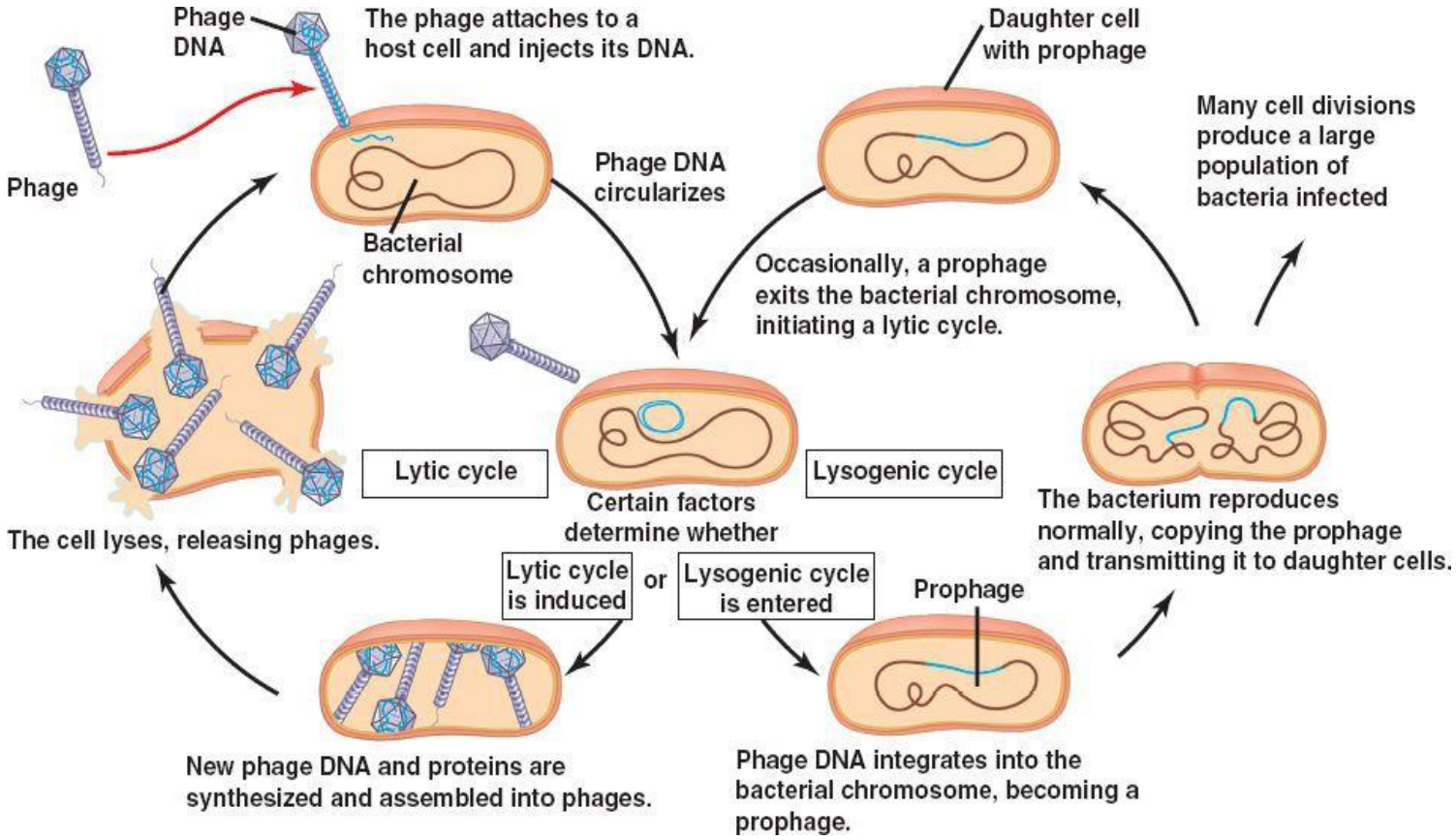


Figure 1.3 Lytic and lysogenic life cycle. This schematic diagram illustrates the life cycle a temperate phage can exhibit. A lytic phage can only infect to replicate (Image taken from (Reese, 2011)).

1.6.3 Phage tail fibre

Siphoviridae phages have long flexible non-contractile tails, *Podoviridae* phages have short non-contractile tails and *Myoviridae* phages have contractile tails (Fig 1.2b). The size of tail length can vary between phages 140 nm 150 nm and 160 nm in *Burkholderia cepacia* complex phages (Lynch et al, 2010). In a study looking at lytic phage of *Pseudomonas aeruginosa* the tail length varied from 110 nm – 170 nm for phage belonging to the *Siphoviridae* family (Sepulveda-Robles et al, 2012). Lytic *Pseudomonas* phage of the *Myoviridae* family showed to have tails ranging between 10nm - 302nm (Sepulveda-Robles et al, 2012). The length of the tail is determined by the tail tape measure protein (Xu et al, 2004). In long tailed phages, genes are highly conserved and have greater adsorption potentials in coliphage lambda (Schwartz, 1976). Following tail adsorption, the tail spikes attach for infection to proceed and inject phage DNA. The lambda tail is 140nm long and encoded by an operon of 11 genes Z, U, V, G, T, H, M, L, K, I and J (Casjens & Hendrix, 1974; Xu et al, 2004). The promoter is the second rightward *p*R' promoter and the 11 genes are self-promoting see chapter 1.6.5.

1.6.2 Cell adsorption and Bacteriophage infection

The adsorption of phage to the host cell is the first step in infection. This is a protein-protein interaction between the virus and the cell wall. This interaction is highly specific as the virus must be able to deliver its genome into the bacterial cell. There are multiple targets phage utilise to infect bacteria (Shao & Wang, 2008). Bacteriophages target a range of cell surface receptors including but not limited to structural proteins that penetrate down to the peptidoglycan layer, porins, enzymes, substrate receptors and transfer proteins (Rakhuba et al, 2010). Upon infection phage specifically adsorb to recognised cellular epitopes, λ tail protein gpJ (Berkane et al, 2006) uses LamB a maltose inducible outer membrane protein (Scandella & Arber, 1974). Linearised phage DNA is injected through the contractile tail across outer and inner membrane. The T7 phage adsorption to the *E. coli* cell surfaces was described as an irreversible step

implying the adsorption has high efficacy to the cell surface ligand (Kemp et al, 2005). The adsorption of T7 phage to the lipopolysaccharides of *E.coli* is electrostatic (Puck et al, 1951).

Alteration in phage ligands can also prevent phage from binding. OmpA is a structural protein that is fixed in the membrane bound to the peptidoglycan layer, found with a free carboxyl terminus (Vogel & Jahnig, 1986). A study showed *E.coli* K12 OmpA mutants defective of this protein based resistance to phage K3, C21 and MS2 infection (Skurray et al, 1974). The study showed that the *E.coli* K12 OmpA mutant were still sensitive to T1-T7, ϕI, ϕII, ϕ3, ϕW, W31,U3 and H (Skurray et al, 1974). This indicated that the phage capable of infecting, adsorb to the cell via another receptor-ligand interaction. A study showed that *E.coli* phage isolated from sewers TuIa, TuIb and TuII used the major outer membrane protein receptors Ia, Ib and II respectively (Datta et al, 1977).

In *E. coli* OmpC and F are porins. *E.coli* phages Hy2, SS4 and T4 are have been shown to adsorb using OmpC (Yu & Mizushima, 1982). OmpF was shown to be a receptor for T2 phage (Hantke, 1978; Riede et al, 1985).

OmpT is an aspartyl protease enzyme and is shown to act as a receptor for adsorption for T-like phages (Hashemolhosseini et al, 1994).

Phages may also target the bacterial flagella and pili, phage chi is one example where it targets *E.coli* via adsorption to the flagella (Schade et al, 1967). It attaches to the flagella and moves down to the base where the flagella is attached to the surface of the bacteria, this is where it injects its DNA (Schade et al, 1967).

Tail fibres of phage are responsible primarily for adsorption of the phage to its host's ligand or epitope. The ligand is most likely to be on the cell surface but phages have been identified to target bacterial flagella. An example of this is bacteriophage χ where it attaches to the filament of a *Salmonella typhi* flagellum (Meynell, 1961). The phage targets the H antigen and the flagellum must be active for the phage infection to take place. Often the phages use the flagella to attach and move along the flagellum to the

base where the actual infection takes place (Schade et al, 1967). This particular study suggested that phage χ has adapted and evolved to only infect motile *Salmonella typhi* strains. Phage χ has also been shown to infect *Serratia* and *E. coli* strains (Schade et al, 1967).

1.6.4 Phage capsid, packaging and phage gDNA ejection

Phages of the *Caudovirales* order tend to have a symmetrical icosahedral capsid, constructed from proteins forming a closed head unit (Figure 1.4B) (Casjens et al, 1985). The capsid proteins are arranged around scaffold proteins (Caspar & Klug, 1962). Figure 1.4C shows a structural illustration of a phi29 scaffold protein. The minor capsid genes encode the proteins that are responsible for connecting the head to the tail (Casjens et al, 1985).This is termed a procapsid as it still needs to be packaged for the head to become mature (Lee et al, 2004). Phage DNA replication can create concatemers arrayed in head-tail axial and terminase resolves the issue by cleaving the DNA once the head is full and is not sequence specific, this is the case for P22, SPP1 and T4 (Aksyuk & Rossmann, 2011; Black, 1989; Black & Rao, 2012). In λ the phage capsid enters the maturation step by the addition of linearised dsDNA cut by terminase at the cos sites (sequence specific), located upstream of the capsid and terminase genes (Figure 1.6). Terminase is also involved in dsDNA packaging in the phage capsid, figure 1.4C, shows small and large terminase proteins of T4 and Sf6 (Black, 1989). The phage DNA enters the capsid through the portal protein, as a response to DNA packaging. In T4, Mu, N4 and Phi29 phage, which do not have concatemeric DNA this event causes a conformational change in the portal protein triggering the terminase to cut the DNA (Black, 1989; Lander et al, 2006; Valpuesta & Carrascosa, 1994). The T4 DNA is packaged longitudinally whereas in T7 it is horizontally packaged (Earnshaw et al, 1978). Figure 1.4A, shows the models of phage DNA packed in various phage heads. The DNA is packaged with enough potential energy in the form of supercoiling the DNA into the capsid. This stored energy is later used by the phage to force the phage gDNA injection. The DNA is packaged with a 12° rotation per 2 base pair this coupled with a turn created

by the packaging motor can generate the energy needed for injection (Spakowitz & Wang, 2005). The scaffold proteins leave the capsid during the maturation step of the head and are either recycled for use in creating another capsid, seen in the maturation step of P22 (King & Casjens, 1974), or in the case of λ the scaffold proteins are proteolytically cleaved (Aksyuk & Rossmann, 2011; Casjens et al, 1985). Figure 1.4 A, illustrates the scaffolding process and depicts how phage DNA is inserted into the procapsid. A study showed this by increasing the osmotic pressure to inhibit lambda phage DNA injection and found that the pressure had to be several times greater to prevent DNA injection (Evilevitch et al, 2003).This implies that the lambda phage capsid with DNA has greater pressure than normal osmotic pressure to allow DNA injection.

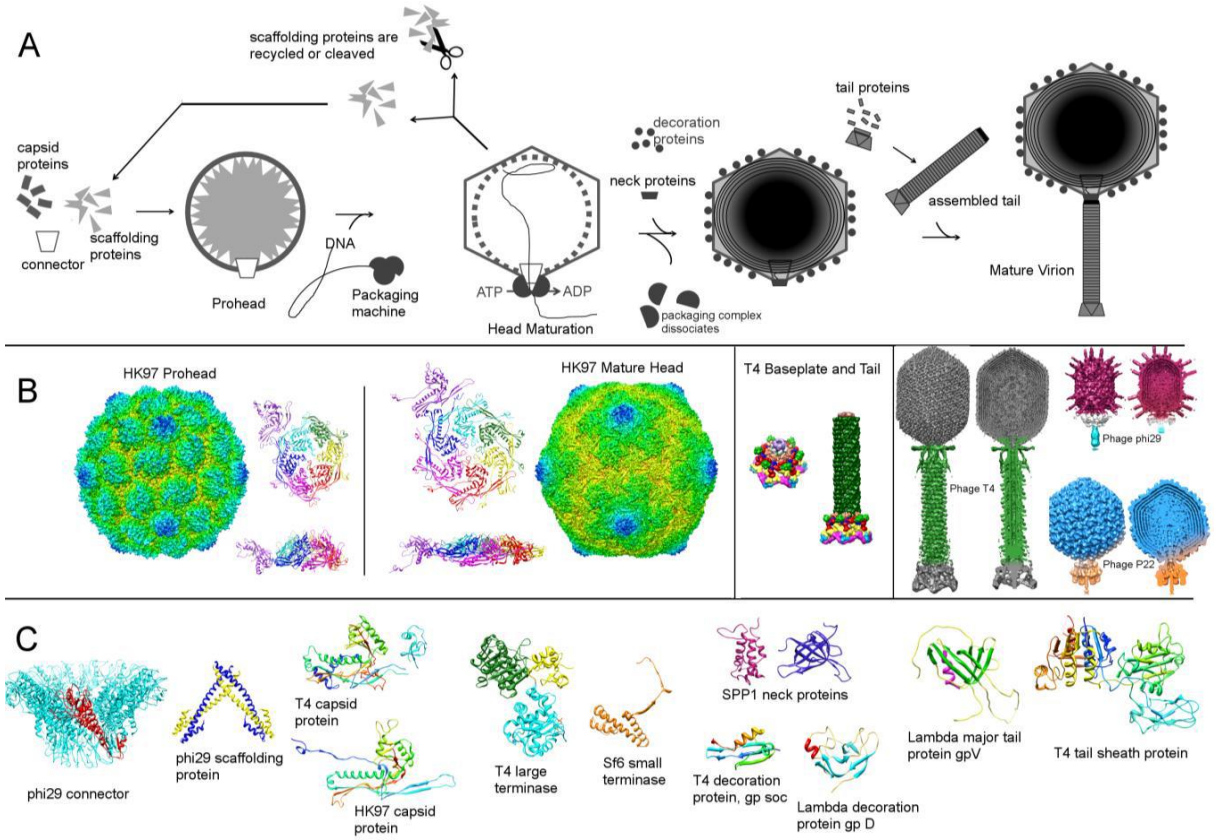


Figure 1.4 dsDNA tailed phage assembly. A) Diagram of phage structural assembly. B) The left panel shows the prohead and mature head of HK97, the middle panel shows the T4 baseplate and tail. The right panel shows T4, P22 and phi29 as mature phages. C) Several structural proteins involved in phage assembly. (Image modified from (Aksyuk & Rossmann, 2011)).

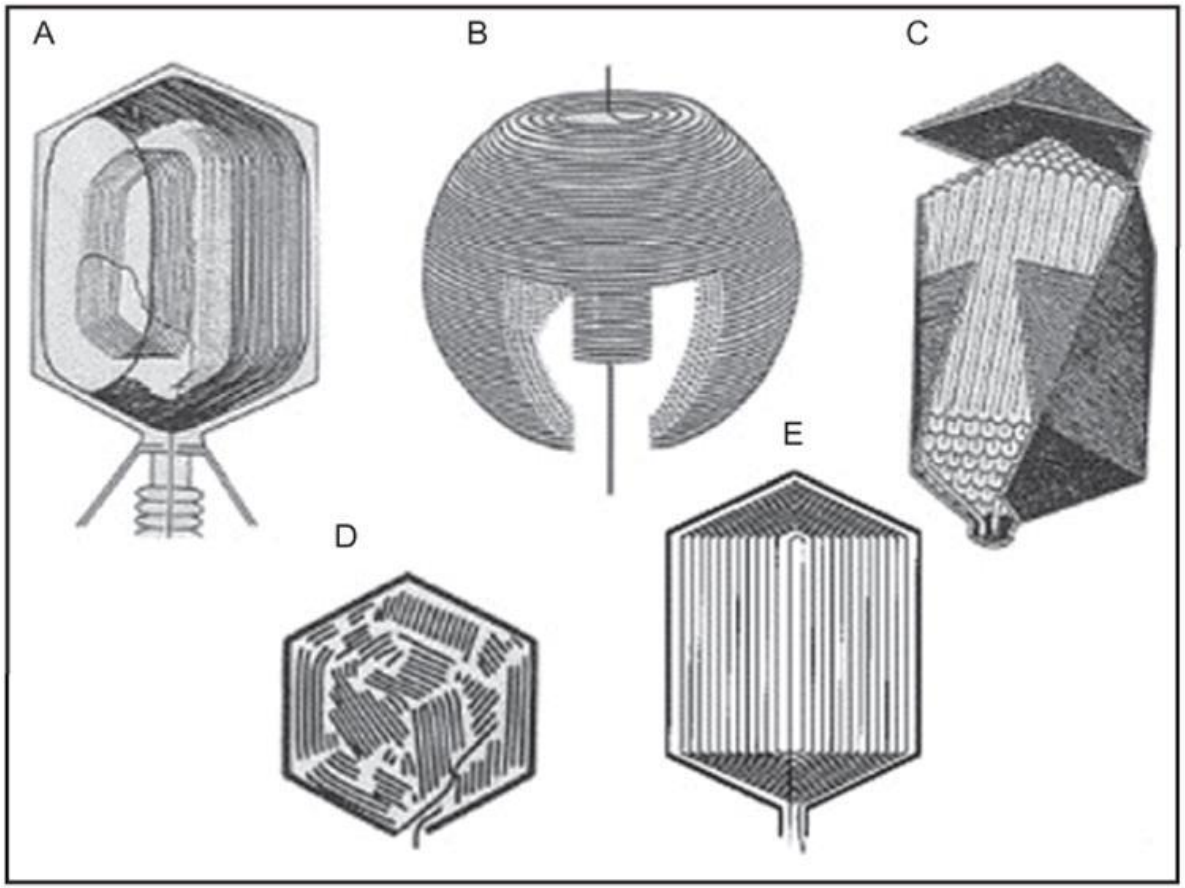


Figure 1.5 Phage DNA packaged models. A) Longitudinal head-tail axial (Earnshaw et al, 1978) B) Spherical C) Liquid crystal D) bends associated with the icosahedral structure E) Both ends of the DNA are at the end of the portal protein (Ray et al, 2010). Image modified from (Black & Rao, 2012) .

1.6.5 Temperate bacteriophages control of life cycle decision through early gene expression on infection

Once the genomic DNA of the temperate phage has been injected into the host bacterial cell there is a rapid decision of whether the lysogenic or lytic life cycle occurs. This is achieved through controlled early gene expression of certain genes encoded by the phage and subversion of the host machinery. Lambdoid phages are a good example of the temperate phages and their life cycle. The genetic regulation of this life cycle switch is discussed further as a model of temperate phage.

The Lambda phage has six major promoters *p*L promoter leftward, *p*R promoter rightward, *p*RE promoter repressor establishment, *p*RM promoter repressor maintenance, *p*I promoter intergration and *p*R', second rightward promoter. Once the linearised phage gene is in the host cell it circularises at the two cos sites using DNA ligase and DNA gyrase (De Wyngaert & Hinkle, 1979), whereby the decision to enter either the lytic or lysogenic life cycle is then made. Temperate phages usually show a high propensity to lysogeny, however, the degree can vary from phage to phage (Davies et al, 2016). Although in either the lytic or lysogenic infection the first priority of phage function is to subvert the host RNA polymerase to transcribe the phage early genes. RNA polymerase binds to four promoters' *p*L, *p*RE, *p*O and *p*R (Ptashne, 2004). The promoter transcription of *p*L encodes N, which is an anti-termination factor "regulator" of early lysogeny gene (Herskowitz & Hagen, 1980). This is the first promoter that is transcribed post infection and cell subversion. Transcription from *p*R encodes cro, this is a repressor regulator of the lytic cycle. It does this by binding to domains of an operator region O_R, preventing premature synthesis of cI the regulatory protein of phage lysogeny and repressor maintenance (Johnson et al, 1978). The expression of phage DNA replication proteins O and P stimulates circular phage DNA replication at location *ori* in the phage DNA (Ptashne, 2004). This then leads to the production of Q anti-termination which is a regulator of late genes such as capsid and tail through modification of the host RNA polymerase.

On injection into the cell if the phage enters the lysogenic life cycle, gene expression from *p*L and *p*RE regulate *cII* and *cIII*, which are necessary for producing cI from promoter *p*RE and promoter *p*int which transcribes for integrase. As the name suggests integrase is needed for the temperate phage to site specifically integrate its DNA in the bacterial genome at location *att*B. The essentiality of integrase in phage insertion into the bacterial chromosome is confirmed by *int* negative mutant (Zissler, 1967). cIII is transcribed from the *p*L promoter and regulates the lysogenic pathway by stabilising cII (Altuvia et al, 1987; Kornitzer et al, 1989), stabilised cII activates cI gene expression (Kornitzer et al, 1991). cI has two promoters *p*RE for establishing lysogeny and *p*RM for maintaining lysogeny (Dodd et al, 2001). The phage *attP* integrates at *attL* and *attR* left and right respectively. To complicate matters if the bacteriophage is to enter the lysogenic life cycle it must ensure cI suppression, as cI is also transcribed from *p*RM, hence cro expression is essential to prevent early expression of cI. In these circumstances cI is transcribed from the promoter *p*RE, which is dependent on cII and cIII. Thus protein expression of cII and *cro* are empirical for life cycle decision (Michalowski & Little, 2005). Table 1.2 illustrates the promoters expressed during the phage replication cycle and the genes their promoters express. Figure 1.6 illustrates these genes in a linear map for bacteriophage lambda, this map is linearised from the cos sites.

Stage	Promoter	Expressed Genes
Transcripts made in lysogenic state	*p*RM	*cI*, $\overbrace{rexA\ and\ rexB}^{lysogenic\ conversion}$
	*p*sieB	$\overbrace{sieB\ and\ esc}^{based\ on\ frameshift}$
	*p*R	*bor and lom*
Immediate-early transcripts	*p*L	*N* gene- a regulatory gene
	*p*R	*Cro* – a regulatory gene
Early Transcripts	*P*o	Ori- origin of replication/ DNA replication
	*p*R	*cII*, $\overbrace{O\ and\ P}^{DNA\ replication}$,ren + rap, $\overbrace{int\ and\ xis}^{site\ specific\ recombination}$, $\overbrace{exo, bet\ and\ gam}^{homologous\ recombination}$ plus other non-essential genes.
Late transcripts	*p*R’	$\overbrace{R, Rz^{'}, and\ Rz}^{lysis}$, $\overbrace{nu1\ and\ A}^{DNA\ packaging/Terminase}$, $\overbrace{B,\ C,\ D\ and\ E}^{Procapsid\ heads}$, $\overbrace{Z, U, V,\ G\ and\ T}^{tail\ shaft}$, $\overbrace{H,\ M,\ L,\ K,\ I\ and\ J}^{tail\ tip}$ and $\overbrace{stf\ and\ tfa}^{side\ tail\ fibres}$
Transcripts made in response to high cII levels	*p*aQ	221
	*p*RE	*Cro, cI* and $\overbrace{rexA\ and\ rexB}^{lysogenic\ conversion}$
	*p*I	$\overbrace{int\ and\ xis}^{site\ specific\ recombination}$ - attP

Table 1.2: Lambda phage promoters, their stages of expression and the genes they express. This table attempts to group the promoter and the genes they express.

Once integrated into the bacterial genome the bacteriophage retains the ability to induce the lytic life cycle. The phage remains in lysogenic state if the λ repressor dimer binds to the O_{R2} operator region repressing *Cro* gene expression. In the lysogenic state O_{R2} and O_{R1} are bound >90% of the time by the repressor and in <10% of the time O_{R3} O_{R2} and O_{R1} are all bound by the repressor (Nelson et al, 2012). The repressor molecules act by blocking RNA polymerase binding to the operator regions. If repressor is bound to O_{R2} *p*R is not expressed and *p*RM is, which promotes cI gene expression. This gene expression is essential for the transcription of *p*int and expression of integrase. Once the phage genome is fully integrated into the bacterial chromosome it is possible to induce the phage to the lytic cycle. Stimulation of cell stress and activation of the RecBCD enzyme starts the repair process of dsDNA and results in SOS induction of phage through auto proteolytic cleavage of cI through stimulation of LexA (Capaldo & Barbour, 1975; Dillingham & Kowalczykowski, 2008). Compounds or environments that have been reported to stimulate this bacterial response and phage induction include: UV (Barnhart et al, 1976; Mattern et al, 1965; Setlow et al, 1973) and antibiotics such as Norfloxacin (Matsushiro et al, 1999). Norfloxacin inhibits DNA gyrase and UV promotes dimerization, thus both initiate the same bacterial SOS response and subsequent RecA pathway activation. RecA activates LexA that is linked to auto proteolytic cleavage of the repressor molecule CI and eventually reducing repressor concentrations in the cell which allows expression of *Cro*, the cro protein bind to O_{R3} and eventually to O_{R2} and O_{R1} preceding lytic growth.

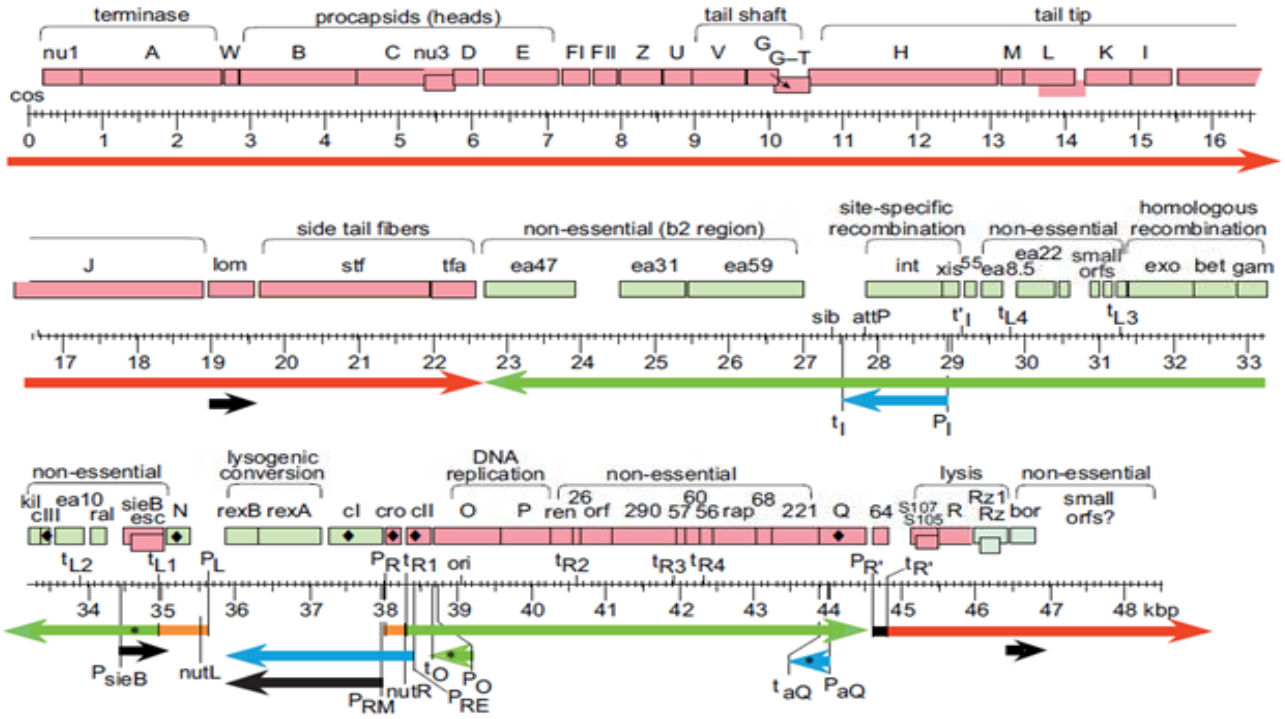


Figure 1.6: Linear genetic map of bacteriophage lambda. Rectangular boxes represent known genes with their names and their functional regions are shown above. Red rectangles denote rightward transcription and green leftward. Rectangles that are offset represent frameshift expression. The black diamond represents regulatory genes. Promoters and terminators are indicated below the genes. The horizontal arrow indicates mRNA's production: black, lysogenic state; orange, immediate-early; green, early transcripts; red, late transcripts and blue mRNA transcripts made in presence of high cII levels. The asterisks represent RNA's with potential regulatory activities. Image and legend taken from (Casjens & Hendrix, 2015)

1.6.6 Cell and community restriction of phage infection

1.6.6.1 Abortive infection

Bacteriophages and their host are constantly adapting and evolving in a host prey relationship, proposed as the red queen hypothesis (Van Valen, 1973). Therefore after the phage DNA is injected into the bacteria, the microbe has evolved mechanisms to evade or control phage infection as they are environmentally outnumbered by phages (Bergh et al, 1989; Chibani-Chennoufi et al, 2004). One method by which bacteria have evolved a strategy to abort infection is by stimulating suicide or by cessation of cellular metabolism (Dy et al, 2014). An example includes a study where no impairment of phage adsorption was seen by a lytic phage yet activity only killed 10 % of the cells indicating other phage mechanisms were in play. Genes that play a role in abortive infection were identified as AbiE and AbiF (Garvey et al, 1995). The AbiE system was shown to halt phage proliferation by causing bacteriostasis (Dy et al, 2014). In phage lambda the RexAB model causes loss of membrane potential, negatively impacting on ATP levels leading to cell death (Snyder, 1995). In the cheese making industry *Lactococcus lactis* is used during the fermentation process and due to focused research to prevent phage infection, 20 Abi systems conveniently named AbiA to AbiZ have been identified (Chopin et al, 2005). This is an altruistic approach by the bacteria as this trait inhibits the phage replicating and infecting other bacteria.

1.6.6.2 Toxin and antitoxin mechanisms

Other strategies that bacteria have evolved to evade bacteriophage infection are toxin/antitoxin systems. These gene regions encode toxins that inhibit bacterial growth and genes encoding antitoxin neutralise the toxins. This is sometimes referred to as the poison-antidote system and was first discovered in a plasmid present in *E. coli* (Ogura & Hiraga, 1983). A review by Magnuson (2007) describes and hypothesises 9 potential toxin/antitoxin functions; junk, stabilisation of genomic parasite, selfish alleles, gene regulation, growth control, persisters, programmed cell arrest and preservation of the

commons, programmed cell death and antiphage (Magnuson, 2007). In terms of bacteriophage, stabilisation of the parasite and antiphage functions are most relevant here. The bacteriophage may encode toxin/antitoxins to maintain their genomes across the host (DeShazer, 2004).

There are three main types of toxin/antitoxin systems. Type I are complementary by base pairing and antitoxin complements the toxin mRNA, which is degraded via RNase III. Hok and sok are toxin/antitoxin systems respectively. The Sok antitoxin is the antisense RNA of the Hok toxin (Fozo et al, 2008). Sok is unstable and decays rapidly while hok is comparatively more stable, thus it was shown that hok is indirectly regulated through mok which aids sok gene expression (Thisted & Gerdes, 1992). An example of type II toxin/antitoxin system is found in the F plasmid of *E. coli.* This system typically codes for the antitoxin protein first due to the genetic architecture and acts to down regulate the toxin's expression. Type II systems vary in the way they exhibit toxicity, the toxin in this example CcdB protein targets DNA gyrase (Bernard & Couturier, 1992). *Erwinia carotova* contains a toxin abortive infection system that can be classed as type III. This method helps to prevent infection by phages (Fineran et al, 2009). In this example two genes of the abortive infection systems the ToxN and ToxI exhibit toxin - antitoxin systems (Fineran et al, 2009). ToxN inhibits bacterial growth and ToxI neutralises this.

Bacteriophage may trigger toxin/antitoxin systems during infection and transcription which in turn would limit phage replication and proliferation, exemplifying the potential of this system (Pecota & Wood, 1996). Bacteriophages have been shown to overcome these systems. For example T4 phages have shown to overcome the activities of Lon proteases in *E.coli* and would thus evade this system (Skorupski et al, 1988).

1.6.6.3 Clustered regularly interspersed palindromic repeat (CRISPR)/CAS

Bacteria have evolved other systems that enable immunity to phage infection including clustered regularly interspersed palindromic repeat (CRISPR). In 1987 Ishino *et al.* found 29 nucleotide repeats with 32 nucleotide spacer sequences but their functions were not known. The CRISPR locus was first characterised by Francisco Mojica and Ruud Jansen and termed CRISPR (Jansen et al, 2002). It was not until 2005 that the spacer sequences were shown to be complementary to bacteriophages (Mojica et al, 2005). This system is an example of adaptive bacterial phage immunity; during subsequent infections, the small spacer fragments aid recognition of phage or plasmids with similar DNA sequences. The CRISPR locus has been found in 84% of archea and 45% of bacterial species of all analysed sequences, illustrating this system is widely found across the bacterial kingdom to prevent phage infection (CRISPRdb, 2016). CRISPR systems have been identified in *Burkholderia* and *Pseudomonas* species. The online CRISPR database shows that one CRISPR structure was found in the *Pseudomonas aeruginosa* LESB58 strain and 3 in *Pseudomonas aeruginosa* M18 (CRISPRdb, 2016). However, no observable CRISPR structures are seen in *Pseudomonas aeruginosa* PAO1 (CRISPRdb, 2016). It can be inferred that this evasion technique evolved due to the “arms race” allowing one strain to survive and be the most successful in the lung of CF patients (LESB58 and M18) and PAO1 which do not incorporate a CRISPR system.

The first component of the CRISPR system is the locus where the genetic memory of previous pathogens is stored. The second component involves CRISPR-associated protein or CAS and is located close to CRISPR locus. CAS is a nuclease protein which cleaves DNA. Three main stages are described for the CRISPR/CAS system (Terns & Terns, 2011). The first is the adaptation stage where CAS proteins take a section of the parasitic DNA and incorporate it to as a novel spacer (Cady et al, 2012). The second stage is termed the expression stage, where the CRISPR locus is copied as RNA (Cady et al, 2012). Ribonucleases cleave this RNA transcript into small CRISPR-

RNAs (crRNA) molecule for each spacer sequence (Brouns et al, 2008). crRNA then forms a complex with the CAS protein and functions very similarly to an antibody molecule as part of the immune system (Brouns et al, 2008; Hale et al, 2009). The final phase is termed interference stage where the crRNA-CAS complex recognises and cleaves the genome of the parasite (Brouns et al, 2008; Cady et al, 2012).

The discovery of CRISPR systems by bacteria inevitably led to the discovery of anti-CRISPR systems in bacteriophage (Mojica et al, 2005). Bacteriophages have been able to counteract the CRISPR bacterial defence systems via nucleotide sequence deletions or mutations within the regions targeted by CRISPR locus (Deveau et al, 2008). A knockout gene study showed that phages infecting *Pseudomonas aeruginosa* with CRISPR systems were only able to with the presence of anti CRISPR genes (Bondy-Denomy et al, 2013) which may be an example of bacteria and phage co-evolution.

1.6.6.4 Restriction Modification (RM) systems

RM systems as the name suggests restrict and modify foreign DNA entering the cell. RM systems are almost globally present in bacteria (Naderer et al, 2002). *Chlamydia* and *Rickettsia* both obligate intracellular parasites are the only known examples that do not have RM systems (Naderer et al, 2002; Stephens et al, 1998). RM systems provide a degree of protection to the host cell against phage infection, an example of this has been shown in lambda phage and *E. coli* host (Arber, 1979). This system relies upon a restriction endonuclease that targets the incoming dsDNA (Naderer et al, 2002). *S*-adenosyl-l-methionine is a methyltransferase endonuclease which modifies the bacterial DNA protecting it against digestion (Naderer et al, 2002).

Bacteriophages have evolved strategies to evade RM systems, the T7 phage encodes for an Orc protein which is complementary to the 24 bp active sites of the restriction endonuclease, thus preventing bacteriophage DNA degradation (Bandyopadhyay et al, 1985). Bacteriophages of *Bacillus subtilis* change their thymine bases with 5-hydroxyme-thyluracil this evades site specific recognition by the RMase

(Kruger & Bickle, 1983), this is also found in T4 infecting *E. coli* (Dam methyltransferase) (Kossykh et al, 1995).

1.6.7 Bacteriophages conversion

Phage integration (prophage) into the bacterial genome and subversion of the host cell metabolism can alter cellular function. Lysogeny has been shown to have an impact on microbial evolution as temperate phage infection can provide a selective advantage for the bacterial host (Sun et al, 2013). The phage can carry toxin genes like *stx* gene in *E. coli* (Plunkett et al, 1999), fitness factor genes like *speA* and *speC* genes that increase disease severity using superantigens (Proft et al, 1999) and phage encoded traits that can change antigen to evade host immune evidently aiding colonisation (Nnalue et al, 1990). In some instances, multiple temperate phages can enter and integrate into the host, an example of this is the Liverpool Epidemic Strain (LES) *Pseudomonas aeruginosa.* The LESB58 strain was discovered in 1996 and was found to be resistant to β lactam antibiotics (Cheng et al, 1996). LESB58 was also shown to infect patients alongside other species of *P. aeruginosa* (McCallum et al, 2002) possibly due to the fact it was more virulent.

When the genome of LESB58 was sequenced it identified five inducible prophage regions (Winstanley et al, 2009). These insertions have increased the genome size of the bacterial host and in this instance its pathogenicity. The addition of extra DNA can be detrimental to the bacterial cell due to a selective cost, although importantly the carriage of other genetic traits encoded by the temperate bacteriophages can aid the bacterial host. Temperate phage encoded traits can help the host adapt to its environment and rapidly evolve, with potential to develop resistance to therapy. An example of this is lambdoid like phages carriage of single stranded recombinases e.g. λRed. This is used commercially as it increases single-stranded recombination 1000 fold (Sharan et al, 2009).

1.6.7.1 Bacteriophage encoding virulence for bacteria as a lysogen

Phages are essentially transposable genetic elements and thus they horizontally disseminate their DNA between their specific host ranges. With this they are able to carry accessory or moron genes that offer a selective advantage for the bacterial host. Temperate phages are able to disseminate antibiotic resistant gene cassettes to make them more resilient to certain antimicrobials although this is not as frequent as previously thought (Enault et al, 2016). Examples of phage-mediated cellular subversion have been well characterised in *Escherichia coli* (Plunkett et al, 1999). Genes such as *bor* of phage λ when expressed has been shown to aid the host bacteria by increasing cellular serum resistance (Barondess & Beckwith, 1995). Lom also encoded by lambda phages aids adherence of *E.coli* to human buccal cells (Pacheco et al, 1997). Potentially this phage-mediated bacterial addition and positive selection also occurs in people with CF and this has been proposed to be linked to temperate bacteriophages (Rolain et al, 2009). The lungs of CF patients provide an ideal niche for emergence of antibiotic resistant strains due to the polymicrobial nature of the lungs, and the increased potential for phage-mediated horizontal gene transfer (Rolain et al, 2009). *P. aeruginosa* is the most common bacteria found in the lungs of CF patients followed closely by *H. Influenzae* (Rogers et al, 2004). Investigating the biology of wild-type phages from clinical isolates, by characterising their biology and genomes, will aid how to model their impact on a bacterial community.

1.7 Current treatments in CF and BR

CF has no cure currently and treatments at present aim to manage or ease symptoms, using chemical or physiotherapeutic methods. For CF and BR both the management of microbial lung infection and the removal of mucus is essential (Ramsey, 1996). Clearing intestinal blockages to aid and provide nutrient absorption is also a necessary management method. CF patients can also suffer from pancreatic blockages, making them insulin dependent; furthermore mucus blockages in the hepatocytes can cause impairments in liver function. CF and BR management methods include thinning

of the mucus to aid its removal thus, mucolytics such as Pulmozyme® and acetylcysteine are sometimes given (Cohen-Cymberknoh et al, 2011). Mucolytics can improve expectoration and the quality of sputum as was seen in a study looking at the role of bromhexine in BR patients (Olivieri et al, 1991). COPD patients would most likely be given flu vaccines to prevent flu related exacerbation. Hypertonic saline can also be administered to draw more water in the lungs, which makes it easier to cough out the mucus (Cohen-Cymberknoh et al, 2011). In severe cases of BR the lower bronchi can be colonised where common antibiotics fail to clear the infection. In such cases the patients may be prescribed with macrolides, azalide and quinolone based antibiotics that have increased penetration abilities (ten Hacken et al, 2007). CF and BR patients will most often be prescribed bronchodilators such as salbutamol, albuterol and salmeterol (Sethi & Cote, 2011). These β2-adrenergic receptor agonists are used for bronchospasm as they help dilate the airways of the lungs (Sestini et al, 2002).

In CF, oral pancreatic enzymes and CFTR potentiators like ivacaftor or kalydeco are prescribed to specific CF patients that have the G551D mutation (Davis et al, 2012). To manage lung infections antibiotics are given often by the aid of a nebuliser, some of the most common are tobramycin (given against *Pseudomonas aeruginosa*), Azithromycin and Ciprofloxacin (Cohen-Cymberknoh et al, 2011). Often a cocktail of antibiotics is used for a prolonged period of time to clear bacterial infection. Antibiotics are used to reduce bacterial burden, both as a therapeutic preventative for exacerbation and for treatment during acute exacerbations. The antibiotics used to clear the infection can be determined by the predominant bacterial infection. Sub-inhibitory concentrations of fluoroquionolones this drug may drive phage/bacterial evolution in the host as cell death may be linked to phage induction and release, driving horizontal gene transfer.

In order to treat severe pulmonary hypotension in BR, Adrenaline (Adr) and Noradrenaline (NAdr) are given to maintain cardiac output. Preoperative nebulized Salbutamol (SAL) is given to prevent bronchospasms (Yim et al, 2002). Both Adr and NAdr are catecholamine, compared to Norfloxacin (NFLX) which initiates SOS response

via the RecABCD bacterial stress response pathway. NFLX is a synthetic fluoroquinolone and is used in this research as a primary temperate phage inducer (Matsushiro et al, 1999). Conversely Adr and NAdr target the QseC cascade which mediates the SOS response and have been shown to induce the expression of phage genes in *E. coli* (Hughes et al, 2009). NFLX has been shown to induce phage in their host and thus has been used to induce out the phage in CF and BR Pa backgrounds in vitro (Matsushiro et al, 1999; McDonald et al, 2010).

1.8 Phages as an antimicrobial therapy

Twort and d'Herelle independently discovered phages in the early 1900's (see chapter 1.6 for further detail) and in 1928 the antibiotic penicillin was discovered by Fleming (Fleming, 2001). There were multiple phage based therapeutics available to treat infections around the time penicillin was discovered (Drulis-Kawa et al, 2015). However, phages were not well understood and for this reason phage treatments failed to have repeatable results giving them a poor impression amongst physicians (Housby & Mann, 2009). In 1940 Chain and Florey developed penicillin to be used as a drug agent and during WWII the drug was in mass production by the drug industries (Chain et al, 2005). While the western countries abandoned phages as treatment options, eastern European countries continued in its use and research. Now with the emergence of drug resistant strains research in phage-based therapies is on the rise again (Brussow, 2005; Drulis-Kawa et al, 2014; Kutter et al, 2010).

Lytic bacteriophages are phages that only replicate in a vegetative way, they do not contain the genetic machinery to integrate in the bacterial chromosome. They infect, replicate and lyse the cell, for this reason they are good choices as a highly specific antimicrobial. Importantly, these phages have evolved to lyse cells and thus carry proteins of interest that may aid therapy, which could be useful as individual therapy options. Lytic bacteriophages have enzymes called depolymerases, responsible for the degradation of the bacterial envelope during infection, and lysis cassette genes

associated with holin, spanins and endolysins (Drulis-Kawa et al, 2015). *E.coli* K1, K95 and *Pseudomonas aeruginosa* bacteriophage depolymerases specifically target their host as the phage adsorbs to its receptors (Castillo & Bartell, 1976; Nimmich, 1994; Pelkonen et al, 1992). Phages like P22 can hydrolyse their host outer lipopolysaccharide membrane with just tail spike interactions (Andres et al, 2010).

Topical use of lytic phage cocktails in treating diabetic wounds have widely been used and results have been promising for patients, one study looked at animal models where a diabetic foot infection with MRSA being treated with lytic bacteriophage and linezolid yielded as an effective treatment (Chhibber et al, 2013). A problem arises when trying to administer the phage internally as gastric enzymes through oral route would degrade as they are structurally made up of proteins (Watanabe et al, 2007). It has been shown that in mouse models with *Burkholderia* infections, the administering of phage to the lung via intranasal and inhalation has promising results (Carmody et al, 2010). The majority of phage therapy involves looking at endolysins, as they are the proteins responsible for the bacterial membrane degradation during cell lysis. However, the majority of these involve looking at lysins for extracellular use in lysing the bacterial cell as opposed to their more common intracellular mode of action.

To complicate infectivity, Gram negative bacteria have two membranes, although the inner wall is thinner than that of a Gram positive bacteria, the addition of an outer wall substantially reduces the efficiency of the phage lysin proteins (Lukacik et al, 2012). It is for this reason that infection by phages in a Gram negative population target porins or channels that would aid delivery of the phage gDNA. *Pseudomonas aeruginosa, Burkholderia cepacia* complex and *Haemophilus influenzae* are all Gram negative bacteria. Not surprising then the majority of phage encoded lysins that are being developed are designed for Gram positive bacteria. Endolysins can be used as antimicrobials and also used clinically as a supporting treatment in areas where current phage therapy is growing (Nelson et al, 2012). Rz and Rz1 are lysis cassettes that are encoded by phage to aid cell lysis (Young et al, 2000).

1.9 DNA sequencing technologies and genome assembly

In the last decade with the advancement of DNA sequencing technologies, sequencing is more cost effective and efficient, with an exponential increase in genome sequences being published (Goodwin et al, 2016; Morozova & Marra, 2008; Zhang et al, 2011). A study compared three current sequencing technologies, Ion Torrent's PGM, Pacific Bioscience' RS and Illumina MiSeq against what they called the market leader, Illumina HiSeq, and found all the sequencers generated usable reads (Quail et al, 2012). However, the accuracy, usability and cost of the platforms varied. The accuracy of the Illumina platforms was shown to be the best at mostly > Q30 accuracy. This means that there is a probability that 1 in 1000 bases is called incorrectly giving 99.9% accuracy on a base call. PacBio RS scored the worst with less than < Q10 (< 90%) (Quail et al, 2012). The required amount of DNA for an Illumina platform is as little as 50ng compared to 1 µg for PacBio RS (Quail et al, 2012). In 1995 Sanger introduced a capillary sequencing machine and was responsible for completing the human genome project in 2001(Collins et al, 2003). The Sanger 3730xl has very high read quality (99.999%) and long read lengths (400 – 900 bp). The major drawbacks Sanger sequencing presents today is high cost and low throughput (Liu et al, 2012). The Illumina MiSeq generates data that is accurate for GC-rich genomes (Scott & Ely, 2015). The Illumina MiSeq has a low run time when compared to the other sequencers, it generates a reasonable sequence yield (~ 15 Gb) and using its short reads allows to build full-length *de novo* sequences making it an ideal choice for multiple bacteriophage sequencing.

The Illumina MiSeq uses sequence by synthesis technology. The three main steps involved are clustering, sequencing/imaging and data analysis. During the clustering process each DNA fragment is isothermally (constant temperature) amplified. The flow cell where all the sequencing takes place is filled with two different oligonucleotides. The DNA fragment has an adapter added which complements the oligonucleotides. Polymerisation of the DNA fragment creates a complementary strand. This double strand fragment is denatured and washed away. The DNA strands then

undergo bridge amplification this is where the DNA strand bends over to the second oligonucleotide mentioned earlier and polymerisation of the strand generates a complementary strand. Then double stranded DNA bridge is then denatured, leaving a forward and reverse strand. This process continues until amplification of clusters for all the DNA fragments are generated. Post bridge amplification all reverse strands are cleaved and washed off, leaving forward strands. At this point the 3' end is blocked, preventing any unwanted priming.

The next step involves sequencing; here a sequence primer is attached to the DNA templates. Next nucleotides which are fluorescently tagged compete to be incorporated. The fluorescent tags are excited by a light source and the results registered. The read products are then washed away. Next the indexed read one primers are read associated with the strand and the read product washed. The same process occurs for the index two primers. After polymerisation of the second oligonucleotides form a double strand bridge which is then linearised and the 3' end blocked. The forward strand is cleaved and washed off. Indexed read two primers are introduced and read until read length is achieved and then washed away (Figure 1.7).

The final part involves data analysis; this is where the sequences are pooled and separated based on their indexes that were introduced during sample preparation. Reads that have similar base calls are locally clustered. The forward and reverse reads are paired thus giving paired end contigs in the form of read 1 and read 2 (Illumina, 2016).

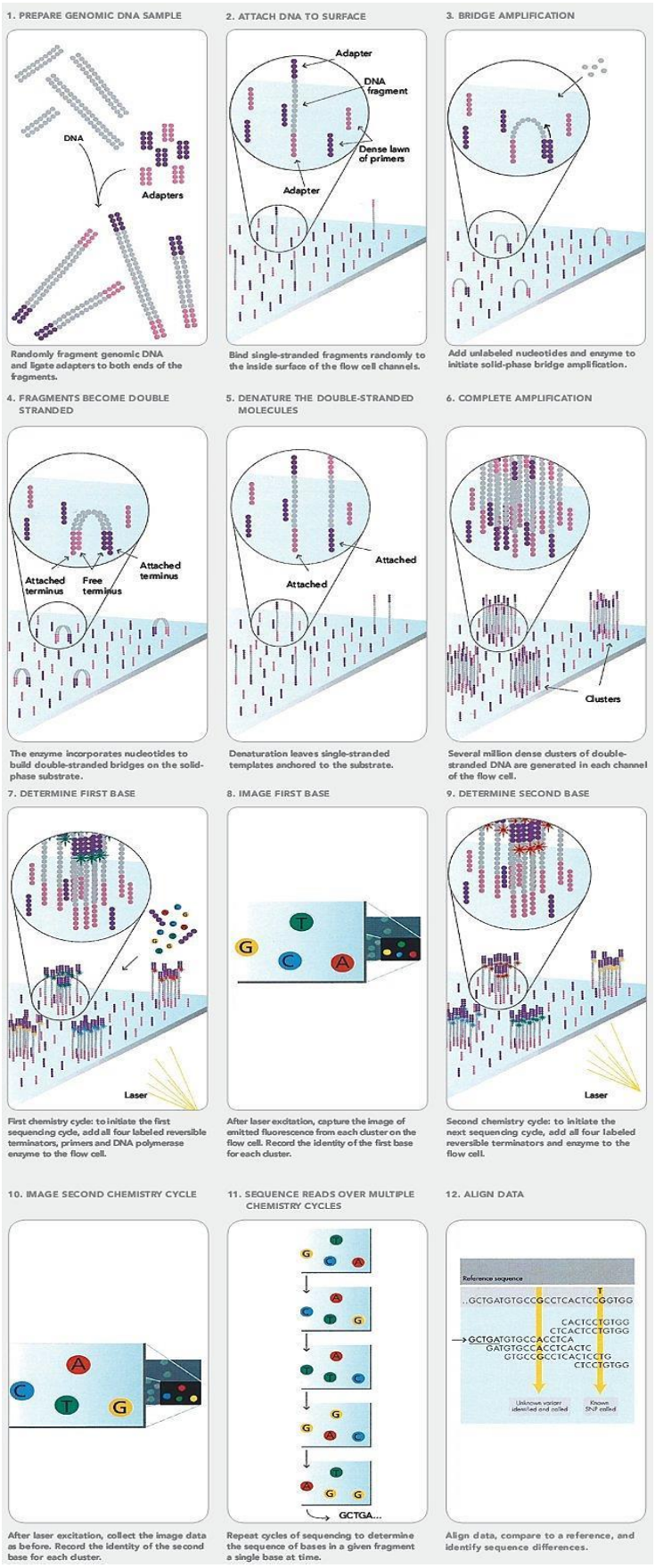


Figure 1.7 Illumina sequencing. This diagram illustrates the workflow of the illumnia next generation sequencer (Slideplayer.inc, 2016).

1.10 Genome annotation

Genome annotation and curating is important before comparisons between genomes can be made. In order to annotate genomes one must predict the open reading frames (ORF's) or coding sequences (CDS's), these ORFs can code for functional proteins (Kropinski et al, 2009). Assigning potential functions to genes is crucial in understanding the phage, and highlights any genes of particular importance. There are many methods to predict ORF's and all automated methods have their advantages and disadvantages. Arguably the most readily used ORF predictors include Glimmer (Delcher et al, 2007), GenemarkHMM (Lukashin & Borodovsky, 1998) and Prodigal (Hyatt et al, 2010). These tools look at all six ORFs, HMM, start and stop codons and Shine-Dalgarno sequence to predict the ORFs. The Shine-Dalgarno consensus sequence is AGGAGG and is found 8 bases upstream of the start codon AUG. The sequence is a ribosomal binding site for mRNA and thus is a powerful predictor for a possible gene. This is because a ribosome protein would attach to the mRNA at the ribosomal binding site and begin translation. Regardless of all these parameters, the fact remains that the ORFs predicted are still putative genes. Glimmer and GenemarkHMM do a good job at predicting ORFs. However, Glimmer and GenemarkHMM still have false positives and incorrect start codon identified for the ORFs. Glimmer is better at calling long ORFs and predicting the start codon when compared to GenemarkHMM.

Prokka is a software package developed by the Victorian Bioinformatics Consortium the same developers of VelvetOptimiser. Prokka v1.11 utilises three main tools to rapidly annotate; Aragorn, Infernal and Prodigal (Seemann, 2014). Prodigal as mentioned before looks for ORFs, Aragorn searches for tRNAs and infernal ("INFERence of RNA ALignment") searches DNA sequence for RNA structure and sequence similarity. Infernal looks at conserved secondary structure rather than the DNA sequence (Nawrocki & Eddy, 2013). Using Prokka as a starting point and for rapid annotation is an ideal choice. Pairing this with manual curation on Artemis to search the ORFs against online databases, this study aims to annotate its temperate phages in this method.

1.11 Aims of this project

This research aims to characterise the impact of temperate phages on their hosts between similar pathophysiological diseases of the lung that are described similarly by bacterial community analysis. The aim is to characterise phages as markers of evolution by looking at their genetic subversion of their bacterial hosts to begin to understand at a phenotypic and genetic level their evolution alongside their bacterial host. This study aimed to look at *Pseudomonas* and *Burkholderia* isolates and characterise the temperate phages from bacterial species associated with these chronic lung conditions with view that data may aid modelling of phage infection, bacterial colonisation and evolution of the bacteria in these microenvironments. As well as their impact on the pathogenicity of the host that will attenuate clinical intervention. This study aims to genomically characterise the induced temperate phages and annotate their genomes.

2. General Materials and Methods

2.1 Materials and growth media constituents

Sterilisation

All the reagents and glassware used were autoclaved at 121 °C for 20 minutes' cycle at a pressure of 15 psi.

Media

Media	**Ingredients**
Bottom LB agar plates (BA);	12.5g (2.5% w/v) of LB broth (Sigma Aldrich, Gillingham, UK) and 7.5g (1.5% w/v) (Lab M, Heywood, UK) of phage agar in 500ml of distilled water.
Blood agar	20g (4% w/v) of Columbia agar (Sigma Aldrich, Gillingham, UK) in 500mL of distilled water and autoclave, allow to cool in 50-55°C water bath, add 5% sterile defibrinated blood (25mL) (TCS Biosciences, Buckingham, UK),
1M $CaCl_2$	147.02g dissolved in 800mL , make up to 1L
Soft (top) agar (SA)	5g (2.5% w/v) of LB broth and 0.8g (0.4% w/v) of phage agar, 0.01M calcium chloride before having 200ml of distilled water added.

Norfloxacin (NFLX) (Sigma Aldrich, Gillingham, UK)	1mg.ml stocks, in water, add a few drops of 1 M NaOH to alter pH so drug dissolves into solution
Phage buffer/Growth media LB Broth, plus 0.01 M $CaCl_2$	12.5 g (2.5% w/v) of LB broth + 5mL of 1M $CaCl_2$ stock in 495mL of distilled water

Table 2.1 List of Media and their constituents

2.1.1 *Pseudomonas aeruginosa* bacterial growth media

All *P. aeruginosa* (Pa) strains were grown at 37 ^{o}C unless stated otherwise either in liquid or on solid Luria Bertani media (LB) (Sigma Aldrich, Gillingham, UK) + 0.1 M $CaCl_2$ (Sigma Aldrich, Gillingham, UK) and high clarity agar (Lab M, Heywood, UK) used for soft agar overlay (0.4%) or bottom agar (1.5%).

2.1.2 *Burkholderia cepacia* complex bacterial growth media

Burkholderia cepacia complex (Bcc) strains are slow growing, even more so from freezer stocks, therefore Bcc strains were grown on 5 % blood for first passage. BCC will grow in 30 °C for 2 days when raised from – 80 °C. For Quality control 2 blood plates with no inoculum were cultured alongside plates to test for contamination.

2.1.3 Storage and maintenance of bacterial cells

All containment level 2 strains (Pa and Bcc) were stored in Glycerol Skimmed Milk (GSM) at -80 ° C. GSM was prepared using Glycerol 20mL (ReAgents, Runcorn, UK) 180mL dH_2O (pH 5.6) 6g TSB (Lab M, Heywood, UK) 4g SMP (Tesco, Newcastle upon Tyne, UK) 1g D-Glucose (Fisher Scientific, Loughborough, UK)

2.2. Bacterial Strains

All bacterial strains were obtained from the Freeman Hospital at Newcastle upon Tyne, NHS trust. The patients' were selected, their sample taken, typed and stored by members of the trust. The patient numbers for Cystic Fibrosis *Pseudomonas aeruginosa* (Pa) and BR Pa are described in Table 2.2.1 and Table 2.2.2. The sample number for Cystic Fibrosis *Burkholdria cepacia* complex (Bcc) strains are described in Table 2.2.3. Ethical approval was granted (REC reference: 12/NE/0248).

CF store ref	VNTR Profiles	Strain type	Adult-Paed	Date of isolation	Age at time of isolation	Mucoid
5	9,2,3,-,3,2,5,3,9	Manchester epidemic strain	A	09/11/10	28 y	N
81	8+,2,3,-,3,2,5,3,9	Manchester epidemic strain	A	08/12/10	49y	Y
136	9,2,3,-,3,2,5,3,8	Manchester epidemic strain	A	02/03/11	24y	N
142	9,2,3,-,3,2,5,3,9	Manchester epidemic strain	A	02/03/11	30y	N
6	11,6,5,3,4,3,5,2,7	LES	A	09/11/10	27y	N
34	11,7,5,4,4,-,5,2,10	LES	A	18/11/10	38y	N
67	11,7,5,4,4,3,5,2,7	LES	A	29/11/10	24y	N
***28**	11,-,5,4,4,3,5,2,7	LES	A	18/11/10	28y	N
16	10,3,5,4,2,2,7,2,9	Leeds epidemic strain	A	10/11/10	25y	Y
145	10,3,5,4,2,2,7,2,9	Leeds epidemic strain	A	04/03/11	44y	N
177	by species specific pcr only	Midlands-1 epidemic strain	A	26/04/11	38y	N
3	11,2,5,3,-,4,12,7,-	unique	A	08/11/10	32y	N
30	11,2,5,3,3,4,12,7,13	same as 30	A	17/11/10	23y	N
24	11,3,1,-,2,2,8,5,11	same as 183	A	15/11/10	39y	N
183	11,3,1,-,2,2,8,5,11	same as 24	A	10/05/11	39y	Y
42	8,2,IS,3,4,IS,9,2,13	same as121	A	22/11/10	27y	N
121	8,2,-,3,4,IS,9,2,13	same as 42	A	25/02/11	19y	N
47	12,3,5,4,2,1,3,3,9	same as 70	P	24/11/10	16y	N
70	12,3,5,4,2,1,3,3,9	same as 47	P	30/11/10	17y	Y

23	8,8,4+,2,1,1,3,2,15	same as 63	A	15/11/10	18y	N
63	8+,8,4+,2,1,1,3,1,-	same as 23	A	27/11/10	29y	N
65	11,6,2,2,1,3,8,3,12	Clone C	A	27/11/10	27y	N
78	11,6,2,2 1,3,7,3,9	Clone C	A	06/12/10	23y	Y
140	8,3,-,5,2,3,5,2,13	same as 208 & 214 & BR 53	A	02/03/11	36y	N
208	8,3,-5,2,3,5,2,11	same as 140 & 214 & BR 53	P	08/07/11	15y	Y
214	8,3,4,5,2,3,5,2,8	same as 140 & 208 & BR 53	P	23/09/11	9y	n
165	11,7,2,2,1,3,7,2,-	same as 187	P	24/03/11	1y 8 months	Y
187	11,7,2,2,1,3,7,2,-	same as 165	P	20/05/11	2y	Y
118	11,5,4,4,3,1,3,1,11	Non defined cluster 2	A	25/02/11	31y	N
44	11,4,5,2,2,_,_,2,-	cluster E	A	24/11/10	41y	N
52	11,2,3,4,2,2,7,3,10	unique	A	26/11/10	25y	Y
53	11,4,5,2,2,3,-,2,8	unique	P	25/11/10	4y	Y
54	11,4,5,2,2,-,-,2,14	unique	A	25/11/10	17y	Y
55	11,2,5,3,3,5,9,7,-	unique	A	25/11/10	25y	N
57	12,2+,5,5,2,3,8,2,12	unique	A	25/11/10	28y	N
60	12,8,4,5,3,3,8,2,10	unique	A	26/11/10	28y	N
69	-,8,1,4,-,1,-,4,16	unique	P	30/11/10	10y	N
72	8,2,9,3,4,2,7,2,11	unique	A	30/11/10	27y	N
74	12,2+,5,5,2,3,6,2,10	unique	A	01/12/10	33y	Y

77	11,4,5,2,3,2,-,4,-	same as bronch 206	A	06/12/10	33y	N
79	8,2,5,3,4,2,7,2,10	unique	A	06/12/10	33y	N
124	11,4,IS,54,5,1,3,2,9	unique	P	22/02/11	10y	N
125	11,3,3,4,2,1,12,2,10	unique	A	16/02/11	59y	N
126	12,6,4,-,4,4,3,6,13	matches another local CF	A	16/02/11	29y	N
127	10,6,4,5,3,2,7,3,11	unique	A	18/02/11	29y	N
211	10,3,5,5,4,1,3,7,8	matches another local CF	A	23/09/11	29y	Y
213	12,2,5,3,3,3,8,5,-	matches another non CF	P	16/09/11	14y	Y

Table 2.2.1: Pa samples number and associated clinical data used in this study. Based on if the samples were classed as adult or paediatric were stratified them into these groups for the purpose of the study.

Bronchiectasis store ref	VNTR Profiles	Strain type	Adult-Paed	Bronchiectasis > 10 yrs	Age at time of isolation	Mucoid
52	10,-,5,5,4,1,3,7,7	unique	A	Yes	56y	Mucoid
53	8,3,4,5,2,3,5,2,8	same as CF 140, 208 & 214	A	Yes	88Y	Mucoid
58	11,6,5,3,3,1,8,3,11	unique	A	Yes	44Y	Mucoid
59	11,6,2,4,3,1,7,3,12	unique	A	No	83y	Non mucoid
123	11,8,-,4,3,1,15,3,10	same as 143	A	?	74y	Non mucoid
133	11,4,IS,6,4,7,4,3,14	unique	A	?	82y	Mucoid
136	11,4,4,5,4,2,6,7,15	unique	A	Yes	64y	Mucoid
141	10,2,-,4,6,2,9,4,12	unique	A	Yes	71y	Mucoid
143	11,8,IS,4,3,1,15,3,10	same as 123	A	Yes	68y	?
144	11,5,9,5,2,5,9,4,14	unique	A	Yes	61y	Mucoid
146	12,3,1,2,4,3,5,2,8	unique	A	Yes	72y	Mucoid
150	12,6,1,3,4,4,6,3,8	unique	A	?	56y	Mucoid
152	12,2+,5,5,2,3,7,1,12	unique	A	No	85y	Mucoid

153	11,-,-,4,3,1,8,1,8	unique	A	Y	62y	Mucoid
161	12,8,2,2,4,3,5,1,11	unique	A	Y	65y	Mucoid
177	10,2,1,4,4,3,3,2,11	unique	A	Yes	61y	Non mucoid
178	11,6,2,2,1,3,8,3,8	Clone C	A	Yes	56y	Mucoid
181	11,3,5,2,2,6,7,3,8	unique	A	Yes	64y	Non mucoid
193	12,5,-,4,2,2,7,3,12	unique	A	?	87y	Mucoid
195	11,4,5,2,2,1,8,2,10	unique	A	Yes	75y	Non mucoid
197	12,2,5,5,3,3,8,2,10	unique	A	Yes	74y	Non mucoid
199	11,3,5,5,3,-,8,4,12	unique	A	Yes	27y	Non mucoid
200	11,4,3,5,4,1,3,2,13	same as another non CF	A	Yes	78y	Non mucoid
201	11,2,5,2,4,1,8,2,10	unique	A	Yes	78y	Mucoid
204	8,3,4,5,2,3,5,2,7	Cluster A	A	Yes	70y	Non mucoid
205	11,6,9,-,3,3,7,2,13	same as another local CF	A	Yes	61y	Non mucoid
206	11,4,5,2,3,2,8,4,11	same as CF 77	A	?	59y	Mucoid

208	11,3,7,5,4,2,4,2,11	unique	A	Yes	61y	Non mucoid
213	11,3,5,4,3,1,13,2,12	unique	A	Yes	64y	Non mucoid
222	12,5,1,5,6,2,7,3,11	same as another local CF	A	Yes	33y	Non mucoid
227	11,4,7,6,4,4,3,2,11	unique	A	?	81y	Non mucoid
228	11,2,5,4,4,4,5,3,7	unique	A	Y	60y	Non mucoid
233	not typed		A	No	87y	Mucoid
243	11,6,2,2,1,3,6,2,12	Clone C	A	Yes	60y	Non mucoid
244	11,6,2,2,1,3,7,2,14	Clone C	A	Yes	76y	Non mucoid
285	not typed		A	Yes	68y	Mucoid
293	not typed		A	No	72y	Non mucoid
299	not typed		A	No	80y	Non mucoid
298	not typed		A	No	87y	Non mucoid
313	not typed		A	Y	38y	Mucoid
319	not typed		A	No	60y	Non mucoid
320	not typed		A	No	63y	Non mucoid
322	not typed		A	Y	74y	Non mucoid

326	not typed		A	Y	73y	Mucoid
327	not typed		A	No	81y	Mucoid
331	not typed		A	No	78y	Mucoid
332	No typed		A	?	?	?

Table 2.2.2: Pa Bronchiectasis (BR) samples number and associated clinical data used in this study. Based on if the samples were classed as over 10 years since clinical diagnosis (> 10 years BR) or under 10 years since clinical diagnosis (< 10 years BR) were stratified them into these groups for the purpose of the study.

Sample number	***Burkholderia* species**	**Patient sample retrieved location and during**
BCC-1	*Burkholderia cenocepacia* genomovar IIIB	Lung transplant FH
BCC-2	*Burkholderia cepacia* group 6	Lung transplant FH
BCC-3	*Burkholderia multivorans*	Lung transplant FH
BCC-4	*Burkholderia multivorans*	Lung transplant FH
BCC-5	*Burkholderia cenocepacia* genomovar IIIA	Lung transplant FH
BCC-6	*Burkholderia cenocepacia* genomovar IIIA	Lung transplant FH
BCC-7	*Burkholderia multivorans*	Lung transplant FH
BCC-8	*Burkholderia vietnamiensis*	Lung transplant FH
BCC-9	*Burkholderia multivorans*	Lung transplant FH
BCC-10	*Burkholderia cenocepacia* genomovar IIIA	Lung transplant FH
BCC-11	*Burkholderia multivorans*	Lung transplant FH
BCC-12	*Burkholderia multivorans*	Lung transplant FH
BCC-13	*Burkholderia multivorans*	Lung transplant FH
BCC-14	*Burkholderia cenocepacia* genomovar IIIA	Lung transplant FH
BCC-15	*Burkholderia vietnamiensis*	Lung transplant FH
BCC-16	*Burkholderia vietnamiensis*	Lung transplant FH
BCC-17	*Burkholderia multivorans*	Lung transplant FH
BCC-18	*Burkholderia cenocepacia* genomovar IIIA	Lung transplant FH
BCC-19	*Burkholderia cenocepacia* genomovar IIIA	Lung transplant FH

BCC-20	*Burkholderia cenocepacia* genomovar IIIA	Lung transplant FH
BCC-21	*Burkholderia cenocepacia* genomovar IIIA	Lung transplant FH
BCC-22	*Burkholderia multivorans*	Lung transplant FH
BCC-23	*Burkholderia multivorans*	Lung transplant FH
24	*Burkholderia multivorans*	Lung transplant Queensland
25	*Burkholderia multivorans*	Lung transplant Queensland
26	*Burkholderia multivorans*	RVI CF
27	*Burkholderia multivorans*	RVI CF
28	*Burkholderia multivorans*	Lung transplant FH
29	*Burkholderia multivorans*	RVI CF
30	*Burkholderia cenocepacia* IIIA	RVI CF
31	*Burkholderia multivorans*	Lung transplant FH
32	*Burkholderia multivorans*	RVI CF
33	*Burkholderia multivorans*	Lung transplant FH
34	*Burkholderia multivorans*	CF Southern Ireland
35	*Burkholderia cenocepacia* genomovar IIIB	RVI CF
36	*Burkholderia multivorans*-1	Lung transplant Queensland
37	*Burkholderia multivorans*-2	Lung transplant Queensland
38	*Burkholderia multivorans*	Lung transplant FH
39	*Burkholderia cenocepacia* genomovar IIIA	RVI CF
40	*Burkholderia stabilis*	CF Southern Ireland
41	*Burkholderia multivorans*	Lung transplant FH
42	*Burkholderia seminalis*	CF Glasgow

43	*Burkholderia stabilis*	Waiting for lung transplant FH
44	*Burkholderia contaminans*	CF Southern Ireland
45	*Burkholderia cenocepacia* genomovar IIIA	RVI CF
46	*Burkholderia cenocepacia* genomovar IIIA	RVI CF
47	*Burkholderia cenocepacia* genomovar IIIA	RVI CF

Table 2.2.3: Bcc samples number and Bcc species used in this study. Acronym; RVI = Royal Victoria infirmary at Newcastle, FH = Freeman hospital at Newcastle, CF = Cystic Fibrosis.

2.3 Temperate Phage induction of the lytic life cycle

Temperate bacteriophages were chemically induced using norfloxacin from bacterial isolates (Matsushiro et al., 1999). Overnight cultures (37 ° C, shaking at 200 RPM) were sub cultured 2 % (v/v) in 10 mL + 0.1 M $CaCl_2$ Luria Bertani (LB) Broth (Sigma Aldrich, Gillingham, UK). At mid exponential growth phase (OD_{600nm}) the bacteria were stressed with norfloxacin (1 µg mL^{-1}) (Sigma Aldrich) for 1 hour (37 ° C, 200 RPM). This culture was diluted in LB (1:10) to reduce the effects of norfloxacin and to begin the recovery for 2 hours at 37 ° C. The lysates were filtered through a 0.22 µM filter (Scientific Laboratory Supplies, Hessle, UK) and stored at 4 ° C for < 1 week. A negative control was always used to validate results by negating the addition of the bacteria.

2.4 Bacteriophage infection and bacterial sensitivity

In order to assess the lysates (methods 2.3.1) contained viable temperate phage, they were spot assayed on a lawn of a suitable indicator strain. This was done by growing the Pa sample to mid-exponential growth phase 0.5-0.6 OD_{600} and 100 µL of this culture added to 5 mL 0.4 % (w/v) soft agar (high clarity agar (Lab M Limited, Heywood, UK) and overlaid onto LB agar plates (1.5 % w/v). 10 µl of each phage lysate was spotted onto a lawn of bacterial host and allowed to air dry. The plates were incubated at 37 ° C for 18 hours and the plaque morphologies analysed. If a zone of clearing was seen depicted at lysis, the lysates were diluted to define individual plaques to negate clearing due to pyocins.

2.5. Cross infection; Statistical analysis and PLS-DA modelling

2.5.1 Disease aetiologies groupings for *Pseudomonas aeruginosa*

Based on table 2.2.1 and 2.2.2 the Pa isolates and their associated phages whenever possible were grouped in the following groups. Adult CF (Adult CF, age >16 years, n = 37), Paediatric CF (Paed CF, age < 16 years, n = 10), Adult BR with < 10 years since clinical diagnosis (n = 17; abbreviated to < 10 years BR) and Adult BR with > 10 years since clinical diagnosis (n = 30; abbreviated to > 10 years BR).

2.5.2 Nestedness and connectance: Anova statistical analysis

Nestedness is a degree of network order where a highly nested pattern describes a hierarchy from generalists (broad host range) to specialists (narrow host range) for both species. The binmatnest value is the deviation of the network pattern from an optimal nested pattern of the same dimensions; values are scaled from 0 (greatly nested) to 100 (not the least bit nested).

Bipartite infection networks of Phage and bacteria interactions both within and between CF and BR were assessed (BR phage vs. CF Pa, & BR phage vs. BR Pa, CF phage vs. CF Pa, CF phage vs. BR Pa). Nestedness was calculated using the binary matrix nestedness temperature computer (binmatnest) (Rodríguez-Gironés and Santamaría, 2006) within the 'nestedness' metric from the R package 'bipartite' - v 2.04 (Dormann et al., 2008, Dormann et al., 2009). The binmatnest values given were statistically significant when compared to the related null model analysis. Network images were generated using R package 'bipartite' - v 2.04. Nestedness and binmatnest definitions

are described in the supplementary data as are the PLS-DA plots that are used to model infectivity profile differences between the aetiologies.

2.5.3 Partial least squares Discriminant Analysis (PLS-DA) modelling of the cross infection data:

The cross infection data's analysis was also performed using the SIMCA-P + v 13.0 software (Umetrics, Umea, Sweden). Before analysis, the data was centre scaled and treated qualitatively to remove the effects of weighting against a zero value; here indicated no infection. The data was analysed using group classification to evaluate the relationships between phage infectivity and bacterial host sensitivity, in relation to patient age, disease aetiology, time since diagnosis and bacterial phenotype (methods 2.5.1). The ability to classify each specific in the correct group was assessed by multiple correlation coefficients R^2Y. The prediction power of the model was evaluated by the Q^2 parameter. The confidence of the data was assessed using Hotelling's *T2* tolerance limits (set at 0.95). Outliers that were deemed to be moderate and unable to shift in the model range were subjected to a DModX (Distance to model in the X space) calculation to weigh those outliers that did not fit the model and stray from the normal F-distribution (critical distance 0.05).

2.6. Phage DNA isolation, genome sequence manipulation and quality control

2.6.1 Pa phage lysate sequencing

For the purpose of simplicity the sample number from Table 2.2.1 and 2.2.2 were re numbered to 1-94. Table 2.6.1 shows the associated samples and their new genome number.

Sample names

	1	2	3	4	5	6	7	8	9	10	11	12
A	3	34	57	74	125	177	59	193	213	298	331	144
B	5	42	60	77	126	183	123	195	222	299	332	146
C	6	44	63	78	127	208	143	197	227	313	187	150
D	16	47	65	79	136	213	152	199	233	319	211	181
E	23	52	67	81	140	214	153	200	243	320	58	201
F	24	53	69	118	142	228	161	204	244	322	133	206
G	28	54	70	121	145	52	177	205	285	326	136	
H	30	55	72	124	165	53	178	208	293	327	141	

Sample number

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	93
F	6	14	22	30	38	46	54	62	70	78	86	94
G	7	15	23	31	39	47	55	63	71	79	87	
H	8	16	24	32	40	48	56	64	72	80	88	

KEY

	Adult CF
	Peadiatric CF
	< 10 years BR
	> 10 years BR

Table 2.6.1 Renumbering of Pa samples to 1-94

The above table show the sample number and their associated re number for the genome section. This was done to simplify and to remove occurrence of repeat reference number for CF and BR pa isolate.

2.6.2 Phage DNA isolation

Following the method described in 2.3.1 and after conformation of phage presence in lysate via method 2.4.1, any remaining bacterial chromosomal DNA within the lysate was decreased using 3 rounds of 1 µL of TURBO DNAse and 1 µL of RNAse Cocktail (Life Technologies Limited), the lysate was incubated at 37 °C for 30 min. The DNAse and RNAse were inactivated using heat at 65 °C and 15 mM EDTA final concentration for 10 min. NORGEN Phage DNA Isolation Kits (Geneflow Limited, Lichfield, UK) were used to filter viral DNA, the manufacturer's protocol was followed apart from the final DNA elution step where 50 µL of elution buffer was used and passed through the elution column twice to increase the total yield. The DNA samples were stored at -80 °C until being sequenced. The NORGEN phage DNA isolation kit was selected due to its cleaner (a decrease in bacterial DNA) and increased yield of phage DNA comparable to the Chelex extraction, QIAGEN QIAmp MinElute Viral Spin Kit, and PEG8000 isolation methods.

2.6.3 Genome sequencing

Genome sequencing of the phage samples were carried out using the Illumina Nextera XT (Illumina, Saffron Waldon, UK) library preparation kit. This was used to prepare and multiplex the isolated phage DNA (methods 2.6.2) for next generation sequencing on the Illumina MiSeq platform. A 2 × 250 cycle V2 kit was used for the loading and running of the sample. The phage DNA samples were quantified and if necessary diluted to 0.2 ng μL^{-1} (Qubit 2.0 DS HS DNA Kit, Life Technologies Limited) earlier to standardisation and combining. The Nextera XT was involved in amplification of DNA, addition of

oligonucleotides sites and addition of barcodes for dual indexing. The samples were then cleaned post PCR using ampure bead solution and magnetic separation solution. The samples were all normalised using Illumina based bead solution, this normalised the concentration of the samples. All samples were pooled and loaded onto the MiSeq. The output data had a Q30 score of >75% for the library constructed.

Paired end sequencing reads were provided as FASTQ files (NU-OMICS, Northumbria University at Newcastle, UK).

2.6.4 Randomizing DNA Reads Using Velvet V1.2.10

The Velvet *de novo* genome assembler package shuffleseq.pl was used to randomly shuffle the FASTQ sequences to limit bias. The following commands were used to do this.

```
$ shufflesequences_fastq.pl read1.fastq read2.fastq shufflesequence.fastq
```

The shuffled sequence output file was directly pipelined into the Khmer toolkit.

2.6.5 Khmer Toolkit

The Khmer toolkit was used to remove very low-level bacterial contamination from the viral sequence data. The Khmer histogram clusters low abundance data and poor sequence data that would be linked to any residual bacterial chromosomal DNA where the defining python script could be to select abundant viral k-mer sequencing data (Crusoe et al., 2015). The following commands were used in order to achieve this.

```
$ load-into-counting.py -N 4 -X 4e9 out.kh shufflesequence.fastq
$ abundance-dist.py out.kh shufflesequence.fastq out.hist
```

For each individual sequence file assessed the out.hist files graphed in excel and manually remove the error k-mer peak. The calc-median-distribution.py script (GitHub,

2015) was altered by a colleague Lauren Cowley, accordingly to select out the k-mers associated to the peaks.

The script shown below was used, depending on the number of peaks and the k-mer range the following lines were changed. In line 34 and 34 a contained the lowest and highest values of the peak in this example 1-901, if the peak number was more than one line 13 was duplicated below and outfile1/peak1.seq changed to outfile2/peak2.seq. 34 to 34 a ix lines were duplicated below and the highest and lowest range changed according to peak2.

```
1.  #! /usr/bin/env python2
2.  #
3.  # This file is part of khmer, http://github.com/ged-lab/khmer/, and is
4.  # Copyright (C) Michigan State University, 2009-2013. It is licensed under
5.  # the three-clause BSD license; see doc/LICENSE.txt.
6.  # Contact: khmer-project@idyll.org
7.  #
8.  import sys
9.  import khmer
10. import argparse
11. import os
12. import screed
13. outfile1 = open('peak1.seq', 'w')


14. def main():
15. parser = argparse.ArgumentParser(
16. description="Output k-mer abundance distribution.")


17. parser.add_argument('hashname')
18. parser.add_argument('seqfile')
19. parser.add_argument('histout')


20. args = parser.parse_args()
21. hashfile = args.hashname
22. seqfile = args.seqfile
23. histout = args.histout


24. outfp = open(histout, 'w')


25. print 'hashtable from', hashfile
26. ht = khmer.load_counting_hash(hashfile)


27. hist = (Crusoe et al., 2015)
```

```
for n, record in enumerate(screed.open(seqfile)):
if n > 0 and n % 100000 == 0:
    print '...', n

seq = record.sequence.replace('N', 'A')
header = record.name
quality = record.accuracy
med, _, _ = ht.get_median_count(seq)
if med > 1:
    if med < 901:
        outfile1.write('@')
        outfile1.write(header)
        outfile1.write('\n')
        outfile1.write(seq)
        outfile1.write('\n')
        outfile1.write('+')
        outfile1.write('\n')
        outfile1.write(quality)
        outfile1.write('\n')

hist[med] = hist.get(med, 0) + 1

maxk = max(hist.keys())

for i in range(maxk + 1):
outfp.write('%d %d\n' % (i, hist.get(i, 0)))
outfp.close()

if __name__ == '__main__':
main()
```

The following command was used to instigate the above script in the directory of choice;

$ python calc-median-distribution.py out.kh shuffleSequence.fastq peak.hist

This would generate separate files containing the peak sequences and the peak.hist file was used to create the final k-mer graphs.

2.7 Genome Assemblers

2.7.1 St. Petersburg genome assembler (SPAdes v3.5.0)

SPAdes v3.5.0 was used to assemble all the phages and the following command was used.

$ spades.py -k 21, 33, 55, 77, 99,127 --careful –s peak.fastq -o SPAdes3.5assembly

2.7.2 VelvetOptimiser (v2.2.5)

VelvetOptimiser was used to optimize the K, expected coverage and coverage cutoff parameters for velveth and velvetg. The command used to assemble the phages can be seen below.

$ VelvetOptimiser.pl –f '-shortPaired –fastq shuffle.fastq' –t 2

2.7.3 Iterative De Bruijn Graph *De Novo* Assembler for uneven sequencing depth (IDBA-UD v1.1.1)

This assembler is designed for short read and sequence data that has uneven sequence depth. The ShortSequence.h file within the package was changed from 'KMaxShortSequence = 128' to 250. This assembler only accepts fasta format files and thus has an fq2fa script.

$ fq2fa read.fq read.fa

In order to invoke the assembler, the following command was used.

$ idba-ud –r reads.fa –mink 20 –maxk 100 –steps 15 –o idba.out

2.7.4 Paired-Read Iterative Contig Extension assembler (PriceTI v1.2)

The script used to execute this assembler in order to iteratively extend contig associated with the same phage can be seen below.

```
#!/usr/bin/env bash
for X in 100 99 98 97 96 95
do  PriceTI –fpp Read1.fastq Read2.fastq 400 90 -rqf 90 .95 -icf contigs.fasta 1 1 1 -nc 51 -nco 1 -mol 30 -mpi $X -target 95 3 2 1 -lenf 50 1 -a 36 -o $.price.fasta
done
```

In the above script read1 and read2 corresponded to the sample and the contigs.fasta file contained the contigs that were being subjected to extension.

2.9 Prokka (v1.11) automated annotation tool

The phage genome were annotated using Prokka, the command used,

$ prokka vB_Pea_CF1a.fasta --outdir vB_Pae_CF1a --prefix prokka --genus Pseudomonas --species phage --strain vB_Pae_1CFa --locustag orf.

The above changed for each phage according to the assigned name.

2.10 HMM/PFAM searches using GeneWise2

We created a Pfam database of 40 amino acid HMMs' shown in Table. GeneWise2 algorithms 6:23 and 21:93 were used to search for putative CAD domains. The alignment comparisons of the domains was shown using Jalview2.

2.11 Basic Local Alignment Search Tool nucleotide (BLASTn) virus database standalone setup.

To search for sequence similarity to the viruses database a standalone BLASTn environment was created.

The standalone blast 2.2.27 package was downloaded from NCBI at ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/ along with the fasta file containing all the fasta sequence (all.fna.tar.gz) in the NCBI viruses' database ftp://ftp.ncbi.nlm.nih.gov/genomes/Viruses. The file was used to create an executable database for blastn searches via the command line. The makeblastdb scripts part of the blast package was used to create a "Viruses" nucleotide database within a db folder.

The following default parameters were used in blastn: word_size = 7, gapopen = 5, gapextend = 2, reward = 2, penalty = -3,

Blastn searches were carried out using the following command.

$ blastn –query query_contig.fasta –num_threads 2 –db Viruses –out blastoutput.txt

3. Characterising the interaction between temperate phages and well-characterised *Pseudomonas aeruginosa* isolates originating from Cystic Fibrosis and Bronchiectasis patients.

3.1 Introduction

3.1.1 Induction and cross infection of temperate phages induced from *Pseudomonas aeruginosa*.

Pseudomonas aeruginosa (Pa) is the one of the most common opportunistic bacterial species that colonises the lung in Cystic Fibrosis (CF) and Bronchiectasis (BR) patients (Kubesch et al, 1988; Purcell et al, 2014b; Roberts et al, 1984; Valenza et al, 2008). Pa becomes more prevalent in CF during second to third decade of life overtaking *Staphylococcus aureus* (Baldan et al, 2014). In one study these phenotypically were seen to be non-mucoid Pa isolates (Valenza et al, 2008). A key feature of Pa growth that aids colonisation and decreases the ability to treat effectively is differentiation in growth to biofilm formation (Costerton et al, 1999). Growth as a biofilm is in a complex matrix of extra-polysaccharides and nucleic acids (Whitchurch et al, 2002). Biofilms are less metabolically active (Walters et al, 2003) and a complex structure all aid as a protective barrier that limits accessibility to antimicrobials. CF patients can become sensitive to Pa infection early in life and at some stage become chronically infected (Lyczak et al, 2002). If we compare this to the aetiology of infection in BR patients the disease is linked to an ageing population or lung trauma.

This study focuses on temperate bacteriophages harboured in Pa isolates as they act as transposable genetic elements that can horizontally transfer genetic information as they integrate into the bacterial host genome (Thomas & Nielsen, 2005). When infection is coupled with high rates of infection and recombination there is the possibility of driving heterogeneity that aids bacterial host fitness (Williams, 2013). In comparison to marine viruses where it is estimated that there are 10^{23} phage infections every second (Suttle, 2007). One could predict that the confined and constrained environment of the CF and BR lower lung may give rise to phage infections at a higher rate.

For a bacteriophage to infect a bacterial cell it needs to have a complementary phenotype and genotype that can infect the bacterial cell and evasion of any evolved resistance mechanisms. This interplay between genome difference and host range is difficult to determine. Ceyssens et al. (2011) describes 4 phages LKD16, ϕKF77 and LUZ19 that share between 83-97% similarities at the genome level. Despite this genome similarity their P. aeruginosa host range showed a marked difference.

As described in 1.5.1, host range is determined as a phage fully infecting and replicating inside of a bacterial cell. Initially the phage must site specifically interact with an epitope at the cell surface, in the case of ϕKF77 is a specific O-chain S type lipopolysaccharide and slime polysaccharides (Bartell et al, 1971). ϕKMV requires type IV pili, incidentally associated with twitching. The ponA and pilMNOPQ genes have been shown to be required to infect its host (Chibeu et al, 2009). Pa phages B3 and D3112 adsorb to the host via the pili illustrated by their inability to adsorb to pili mutants (Roncero et al, 1990). D3112 also prevents phage superinfection as it encodes a Tip protein that binds to TFT ATPase which results in loss of the pili (Chung et al, 2014).

Once adsorbed to the cell surface the phage must be able to deliver genomic nucleic acids directly across the outer and inner membrane in Pa. This causes an increase in extracellular K+ ions occurring due to cytoplasmic leakage (Cady et al, 2012). The phage nucleic acid is now subject to host restriction and other cell mechanisms evolved to eliminate incoming transposable genetic elements. These include CRISPR Cas, abortive infection, toxin/antitoxin and restriction modification systems (chapter 1.6.6). Bacteria have also adopted systems including abortive infection and toxin-antitoxin systems to evade detection by the host (chapter 1.6.6).

3.1.2 Phages offering accessory genes

Temperate bacteriophages can deliver accessory or moron genes that may aid bacterial host fitness, offering an advantage and protection to the residing prophage. Horizontal gene transfer to specific loci can help phage adapt as was shown for Mu-like

Pa temperate phages, which was driven by selective pressure of the host (Cazares et al, 2014). Phages may acquire accessory genes through recombination events which can that can give rise to fully functional progeny (Hendrix, 2002; Hendrix, 2003). The addition of these accessory genes via lysogenic phages can sometimes make a bacterial strain more virulent. Cholera caused by *Vibrio cholerae* produces a toxin that is encoded by the ϕCTX (Karaolis et al, 1999; Waldor & Mekalanos, 1996). *Corynebacterium diphtheriae* causes Diphtheria the exotoxin DIP0222 it produces which alters elongation factor 2 inhibiting protein syntheses is encoded by ϕβ (Holmes, 2000; Leong & Murphy, 1985). In *S. aureus*, ϕETA encodes for an exfoliation toxin causing scalded skin syndrome and in babies can result in toxic shock like syndrome (Dajani, 1972; Reichardt et al, 1993). A *S. aureus* phage ϕN315 enhances immune evasion by Sak gene expression which codes for staphylokinase (Sako & Tsuchida, 1983).

Aims: The aim of this work was a combined study to characterise the difference in infectivity of, bacterial sensitivity to temperate phages chemically induced from *Pseudomonas aeruginosa* (Pa). All isolates used in this study are from Cystic Fibrosis (CF) and Bronchiectasis (BR) patients. I specifically targeted phage induced from BR Pa isolates where the meta data is combined and presented with the CF phage cross infection data by Francesca Everest (unpublished data) so that they can be modelled and compared alongside clinical data to contrast between diseases and longevity of disease. A panel of 47 CF Pa and 47 BR Pa bacterial isolates were used that are supported with in-depth clinical data including; the patients age (CF/BR), length of time since diagnosis (BR) if the Pa isolates are mucoid or non-mucoid (CF/BR) and the antibiotic sensitivity of the samples (CF/BR). The temperate phages where chemically induced from each bacterial host by targeting bacterial DNA gyrase, using the fluoroquinolone norfloxacin (NFLX). This places a cellular stress on the Pa isolate and stimulation of the RecABCD pathway and phage induction (Matsushiro et al, 1999). The hypothesis to be tested in this study is that there is a continual selective pressure

between the bacterium and its predating phage. We hypothesise that firstly, there will be a difference in phage infection and bacterial sensitivity profiles between CF and BR patients. Secondly, the study aims to test the hypothesis that CF patients colonised with Pa will have undergone increased rounds of proliferation or exclusion and therefore these higher levels of interactions including recombination will show a determinable difference. This would possibly correlate to a potential increase in adaptation and evolution in the lung.

3.2 Results

Phage lysates were induced from the 94 bacterial isolates using NFLX as described in methods section 2.3.1. It is highly likely based on previous research that there are multiple phages that are inducible within the chromosome of Pa bacterial isolates. Section 1.3 of the introduction discusses the level of phage infection in Pa, which could be termed polylysogeny and therefore it is possible that each of the lysates contain more than one bacteriophage. Each lysate containing a possible mixed-phage collection was horizontally infected against all isolates including the isolate the induction was performed. This study therefore looks at the range and distribution of infection the complement of phages can achieve. However, this study cannot infer which phage is responsible for the infections observed. But importantly it does show the potential within the phage lysate of the spectrum of isolates at least one phage within the lysate can infect. This may have larger ramifications in the lung, as it shows each bacterial isolate may harbour multiple phages, especially if there are core genes that each phage carries that aid a specific selective advantage that is needed in the lung for the bacteria to survive.

3.2.1 Infection profile of inducible bacteriophage

The 94 induced phage lysates were sequentially infected across the whole panel of 94 Pa isolates. Each lysate was spotted onto a lawn of growing bacteria as detailed in methods chapter 2.4.1. As pyocins are associated with stationary phase growth cells, phage induction was carried out during mid-exponential growth phase and if samples were dilute plaques instead of zones of clearing could be found and purified if necessary. In total including the CF cross infections completed by Francesca Everest, over 9,000 infection spots were performed.

Three viral plaque morphologies were determined; lysis, turbid and halo. In appendix 1 the cross infection raw data contains this detail. To categorise each for this

study is as follows; lysis showed complete clearing. Halo was when there was a distinct zone of clearing but a background level of bacterial growth around the outer of the plaque, possibly representative of titre and lysogen formation rather than lytic infection. Turbid infection was termed when a lower level growth could be seen across the zone of clearing, showing either resistant mutants or most probably lysogens. Buffer was also spotted onto plate as a liquid mark control in order to eliminate visual bias.

Figure 3.1 displays a heat map of the raw cross infection data detailed in appendix 1 directed as infection of phage lysate or no infection of phage lysate, respectively coloured as yellow or red. Figure 3.1 also shows the infection profiles grouped based on the disease aetiologies. The CF phages by eye are seen to be more infective across the Pa bacterial isolates represented by the depth in yellow in figure 3.1. The BR phages in comparison show fewer infections. To further discriminate between the different disease sub aetiologies (chapter 2.5.1) it is seen that paediatric CF phage and < 10 years BR exhibited fewer infection. However, correlating the infection profiles to the clinical data shows that some of these strains are variable number tandem repeat (VNTR) typed for identification as similar to one another. The VNTR typing method is used for genetic based bacterial identification. Despite this varying phage infection profiles are seen on these strains. For example, in CF paediatric group the phage from CF 214 exhibits the lowest infection rate (8.5%) in this disease aetiology. This strain is from a 9-year-old CF patient. This strain is marked as similar to the bacterial isolates CF140 (adult), CF 208 (paediatric) and BR 53 (adult). There phage infection rates are 69%, 94% and 86% respectively. This shows that the clinical isolates strain typing does not reflect the phage infection profile. The heat map shows phage isolated from CF 3 and CF 5 bacterial isolates exhibiting 100% infectivity across the Pa cohort. Both from adults CF patient of non-mucoid isolates. CF 5 is from a Manchester epidemic strain. Further clinical data from the CF cohort defining the date the patient became chronically infected with Pa may answer some of these differences.

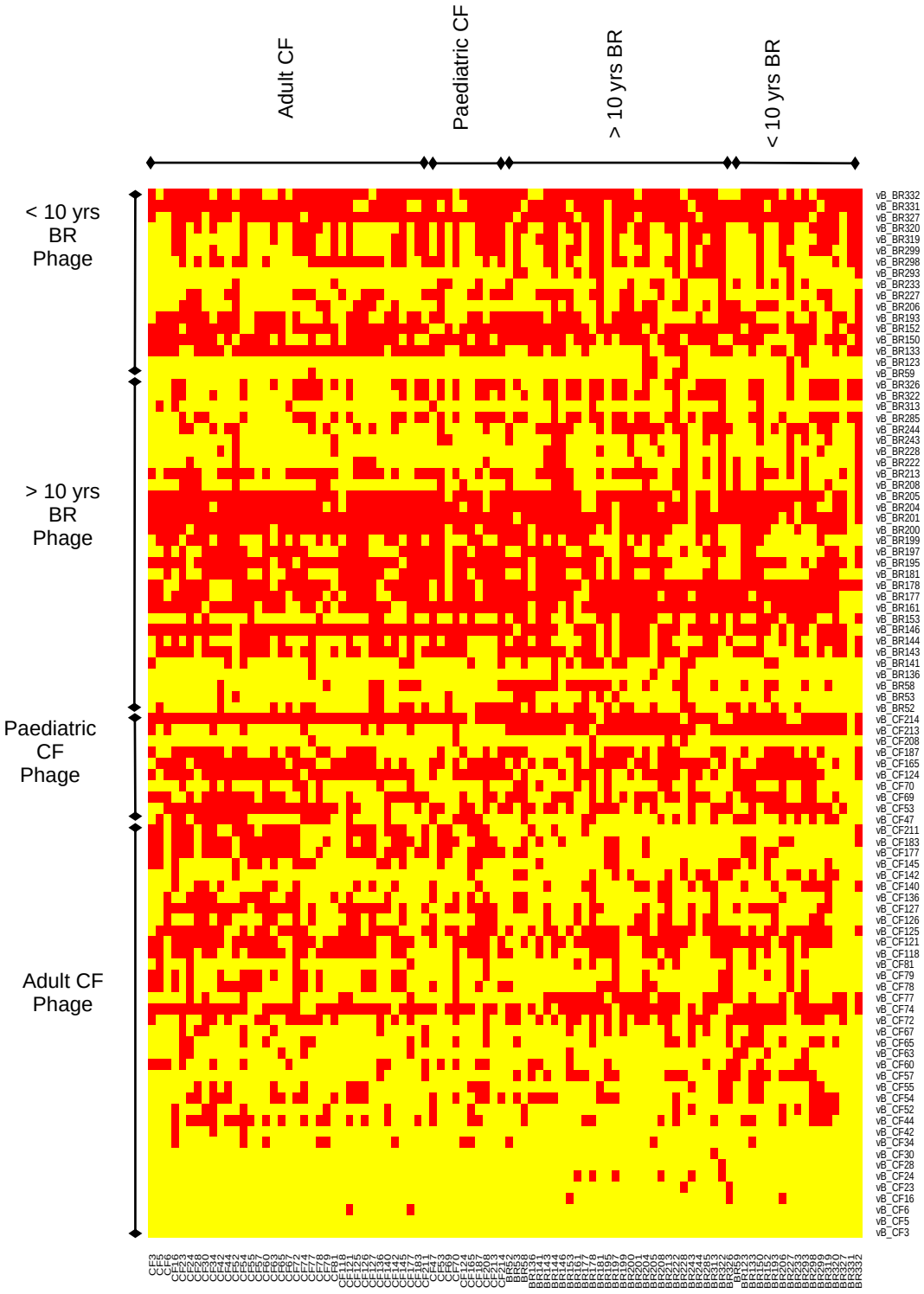


Figure 3.1: Heat map of phage cross infection profile. Heat map generated using R to illustrate the infection profile of cross infection data (Appendix 1); Red represents no phage infection and yellow represents phage infection. The heat map illustrates that the adult CF phage are more infectious than BR phage. The map is based on disease aetiologies described in the methods section 2.5.1.

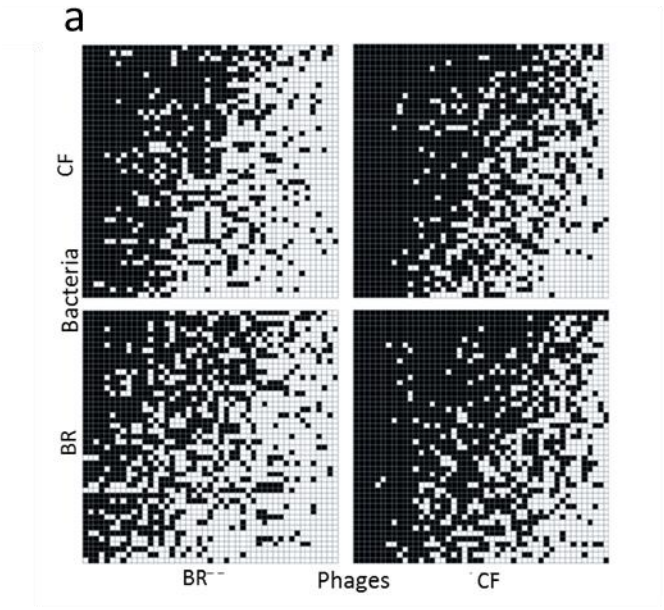
3.2.2 ANOVA nestedness and connectance plot of cross infection data.

The cross infection data was subjected to an ANOVA, bi-partite statistical analysis that was carried out in collaboration with Professor Mike Brockhurst, University of York, York, UK. This approach and method is described in chapter 2.5.2. This analysis adds statistical support to the raw data in appendix 1 and the heat-map figure 3.1.

Figure 3.2.a shows that infection of BR and CF phages to infect the entire Pa cohort. A black square denotes infection. This data suggests that the adult CF phages have a broader infection profile across the Pa cohort based on nestedness and statistically significant manner. The model creates a gradient of the most infective to least infective within the group's quadrant.

Figure 3.2.b shows the data modelled against the mucoid phenotypes (clinical data presented in table 2.2.1 and 2.2.2). Mucoid BR Pa isolates (n = 22) were the only clinical parameter from which any significant difference could be drawn. The panel indicates that BR phages could infect mucoid BR Pa bacterial isolates at a greater rate when compared to non-mucoid BR Pa bacterial isolates. This could reflect the mucus thick physiology of the chronic lung in BR patients, where the phage would infect and propagate.

Figure 3.2.c tabulates the associated statistical data relating to figure 3.2.a and 3.2.b. The ANOVA test helps add statistical based analysis to the results alongside assisting description of the data and support the hypothesis proposed.



c

	Connectance	Nestedness (binmatnest)
(A) Φ CF V CF	0.659	26.9
(A) Φ BR V CF	0.528	28.2
(A) Φ BR V BR	0.486	38.5
(A) Φ CF V BR	0.693	24.9
(B) All Φ BR x Mucoid BR	0.502	28.6
All Φ BR x Non Mucoid BR	0.494	39.9

Figure 3.2: Nestedness and connectance plots of the cross infection data

a) Illustrates the binary nested network for the 94 Phage samples against 94 Pa isolates, ordered by nestedness within quadrants (47 CF Phage vs 47 CF Pa isolates, 47 BR phage vs 47 CF Pa isolates, 47 CF phage vs 47 BR Pa isolates and 47 BR phage vs 47 BR Pa isolates). Each black square represents an infection between one phage lysate (x-axis) and one bacterial strain (y-axis). Grey squares signify no phage

infection. The CF phages were capable of infecting BR Pa isolates more readily than the BR phages could infect CF Pa isolates.

b) Displays the binary nested network for the infection of mucoid BR (table 2.2.2) Pa isolates (n = 22) with the entire cohort of mixed temperate phage communities induced from BR Pa isolates (n = 47)

c) Shows the numerical values for the results shown in 3.2a.

3.2.3 Modelling of cross infection data

The cross infection data appendix 1 and figure 3.1 were subjected to further analysis in order to discriminate between the infection profile data as a whole. Thus cross infection data was extrapolated and modelled using Principle least squares discriminant analysis (PLS-DA) which is a semi-supervised modelling technique, maximising the variation in the sample that allows for trends to be seen. In this instant either phage infection or bacterial sensitivity to phage infection was the parameter variation sought.

Simca – P was used to create PLS-DA. The PLS-DA plots in Fig 3.3 illustrate differences between cross infection data that has been further supported by colouring the points to inform on aetiology of the patient's infection.

Figure 3.3 shows four PLS-DA plots looking at similarities and differences between phages and their bacterial host. It shows variance between phage infectivity and bacterial susceptibility (Panel A-C). The quality of the model is represented by the R2Y value. > 0.5 is considered robust; the plot sees a value of 0.72. Figure 3.3 panel A illustrates the adult CF phages represented in blue are different from paediatric CF phage infection profiles shown in yellow.

The PLS-DA plot shown in figure 3.3 panel B illustrates the difference in infection profile between CF and BR phages. CF phages are depicted as red and BR as blue. The two groups stratify from one another on the x-axis. There is some similarity and cross over where the CF phage infection profiles associated to the bacterial isolate CF214, CF165 and CF213 (table 2.2.1) are seen to be associated with paediatric patients. The fourth sample CF77 which is disparate belongs to a 33-year-old adult CF patient. The R2Y value is 0.77 which indicates there is a significant difference.

Figure 3.3.C illustrates a PLS-DA plot that showing differences now not in phage infectivity but more how sensitive the bacteria is to infection by this phage panel.

Specifically, here between the entirety of the CF and BR Pa bacterial panel. BR isolates are represented as blue and CF isolates as green. The R2Y value for this model is 0.83, when considering > 0.5 is significant this is a strong indication that there is difference seen that is supported by a strong model of the data. The two sensitivity profiles between the CF and BR Pa group apart but two CF Pa isolates are seen to group with the BR Pa isolates. These CF Pa isolates originate from adult CF bacteria. These two CF Pa isolates are CF136 (30-year-old) and CF177 (38-year-old) (table 2.2.1). The clinical data also highlight that the isolates are both epidemic strains from Manchester and the Midlands respectively (table 2.2.1)

Figure 3.3.D is a PLS-DA plot of the sensitivity to phage infection of the CF Pa isolates to phage infection. Paediatric CF Pa isolates are shown in red and adult CF Pa isolates are shown in blue. The phages separate but it is not as clear as 3.3A. The R2Y value for this model is not as robust with 0.66, although still greater than the significant cut-off limit of > 0.5 for the model integrity.

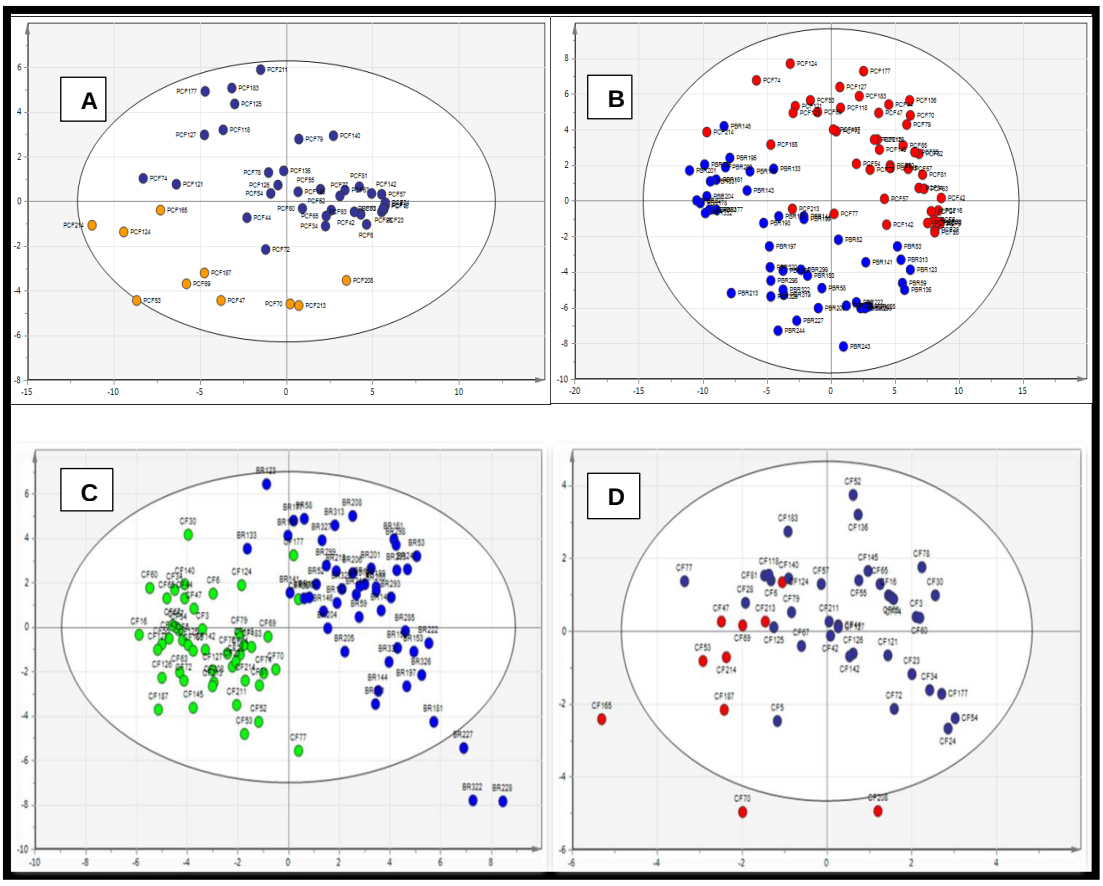


Figure 3.3: Phage infection and bacterial susceptibility PLS-DA Simca – P was used to generate these PLS – DA plots to show differences in phage infectivity and bacterial sensitivity to phage infection, based on these data shown in figure 3.1 and appendix 1, confidence limits are > 80% and integrity is > 0.5. A) Bacterial susceptibility of all the host bacteria to both adult CF phage (blue) and paediatric CF phage (yellow). It shows that there is a difference in bacterial susceptibility to the different phage groups as the stratification in this figure demonstrates (R2Y cumulatively is 77%). B) Shows the bacterial susceptibility of the bacterial hosts to the phage infection from the whole BR cohort (blue) and the whole CF cohort (red). Stratification between the groups is again observed and the CF samples which associate with the BR strains originate from paediatric CF patients (R2Y cumulatively is 72%). C) Phage infectivity differences between the entire CF cohort (green) and the entire BR cohort (blue).

There is clear stratification between phage infectivity profiles, but there is some cross over like in Figure 3.3.B and the relationship of the phage is between the paediatric CF isolates and the BR isolates with < 10 years *P. aeruginosa* infection (R2Y cumulatively is 82.5%). D) Phage infectivity differences seen between the CF phage isolates, adult (blue) and paediatric (red). There is stratification between the different phage lysates but it is not as clear as in Figure 3.3.A (R2Y cumulatively is 66%).

Figure 3.4 models BR bacterial susceptibility to phage based on the length of time the bacterial host has colonised the patient's lung. The most recently diagnosed with Br isolates (yellow) pull away from the chronically colonised strains > 5 years (light blue) > 10 years (royal blue)

Further stratification of the data was done using the PLS-DA plots based on mucoid and non-mucoid plots. In figure 3.5 the PLS-DA plot shows the mucoid isolates as red and the non-mucoid isolates are shown as blue. The BR Phage show no difference between infecting mucoid and non-mucoid strains (figure 3.5 panel a). However, CF phage infect more non-mucoid strains showing a difference (figure 3.5 b)

The PLS-DA presented in Figure 3.5.a focuses on phage infectivity of BR phages infectivity when the data is informed to show the difference in ability of the phages to infect a Pa isolates with a mucoid phenotype. The mucoid variants compared with the non-mucoid variants do not separate as clearly, but it does replicate the findings of the bi-partite modelling seen figure 3.2. The R2Y value is 0.67.

Figure 3.5.b details the variation in the infectivity profiles of the CF phages when they are grouped according to mucoid phenotype. There are two samples in this PLS-DA which are outside of the confidence eclipse (95 %), both are phages induced from mucoid Pa bacterial isolates. Phage from CF54 Pa isolates which is from an adult CF patient (17 years) whilst phage from CF187 is from a paediatric CF patient 2 years of age. The R2Y value is 0.72.

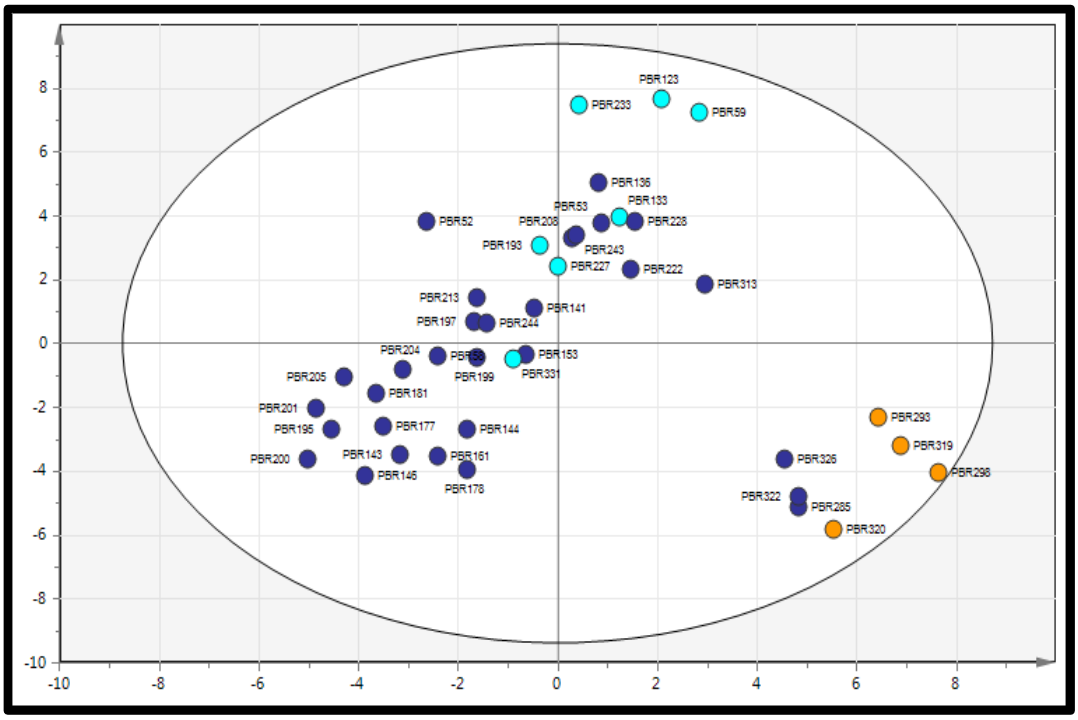


Figure 3.4: Pa difference mapped against disease aetiology. Simca P has been used to generate this PLS DA plot to show the length of time that each BR isolate has been colonised with *P. aeruginosa*. The confidence limits are > 80% and the integrity is > 0.5 and the R2Y value cumulatively is 40%. It shows that with a few exceptions that there is stratification of the most recently colonised hosts (yellow) away from the chronically colonised strains. The patients that have been colonised for over 5 years (light blue) and over 10 years (dark blue) do stratify away from each other but not as discreetly as the under 5 years' colonisation isolate.

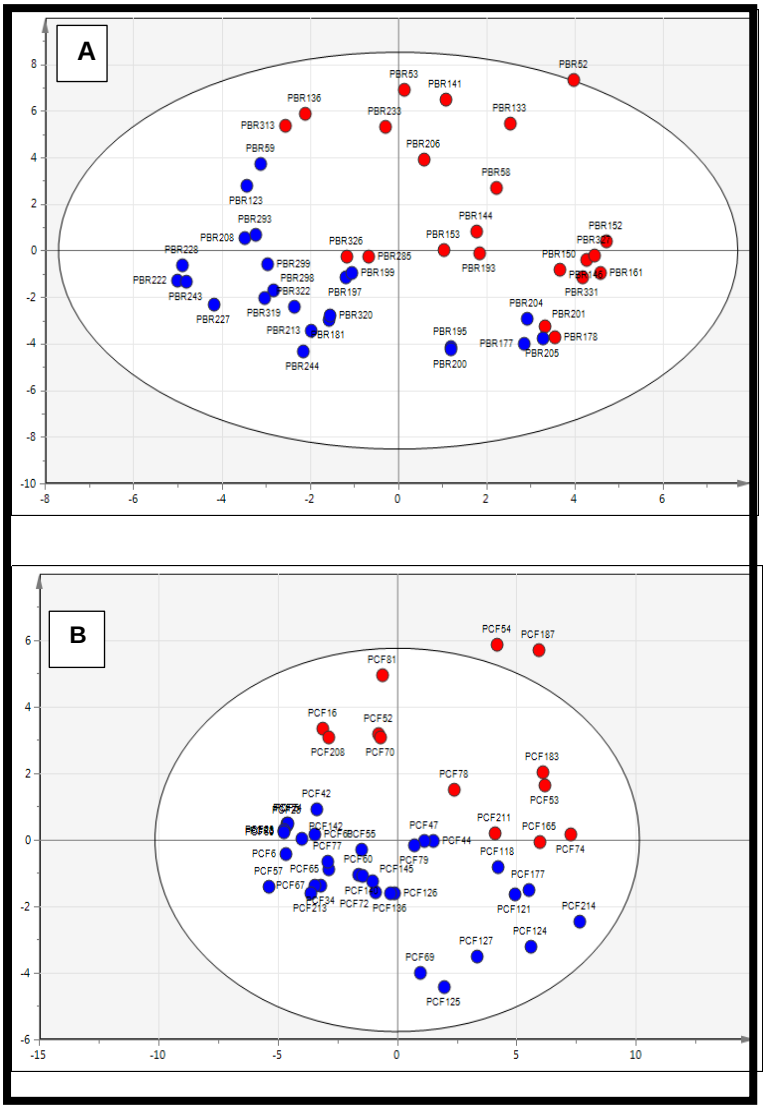
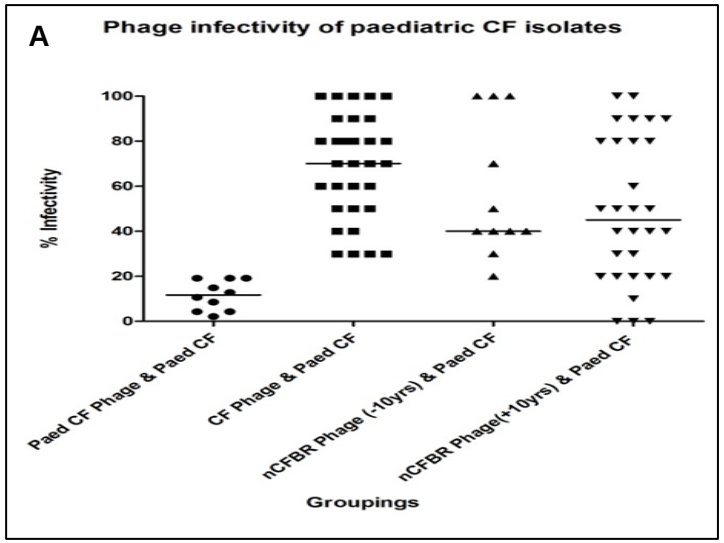
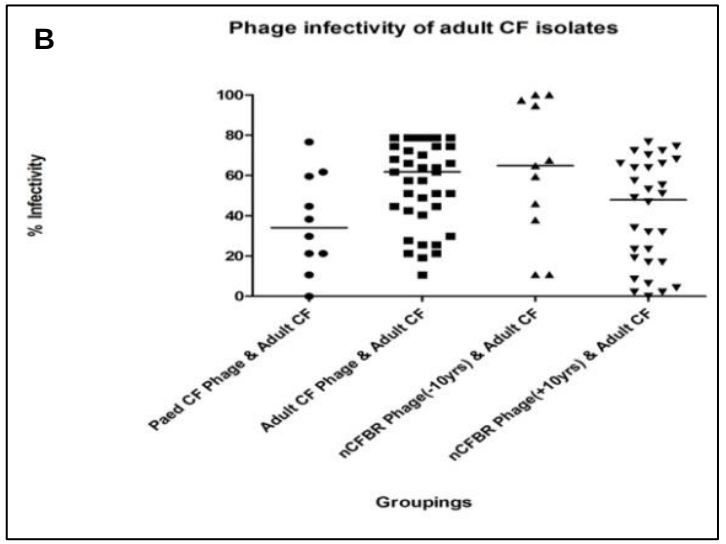


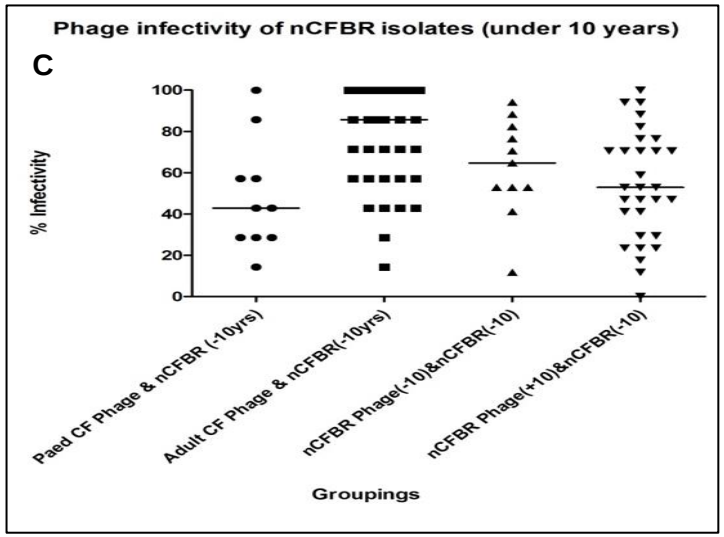
Figure 3.5: Phage infection difference between mucoid and non-mucoid isolates. Simca P has been used to generate PLS DA plots to show the variation in phage infectivity when the host strain phenotypes are mucoid or not, the confidence limits are > 80% and the integrity is > 0.5. A) Phage from BR isolates appear to be able to infect both mucoid (red) and non-mucoid (blue) versions of the host at equivalent infection levels. The stratification is not as distinct as in some of the other PLS DA models but it is still apparent between the two cohorts (R2Y cumulatively is 67%). B) Phage from CF isolates appear to infect more non mucoid hosts (blue) than mucoid (red), this is concluded because the mucoid isolates are more tightly connected on this

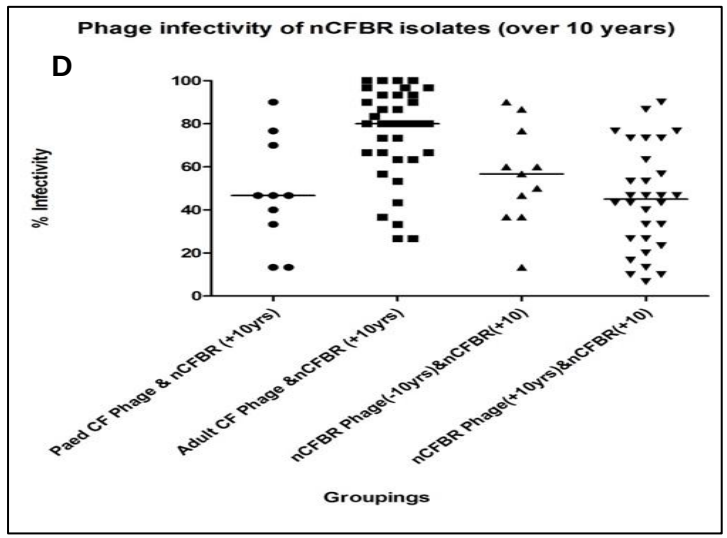
model. There is stratification between the two groups but the separation is not as discrete between the two different host phenotypes (R2Y cumulatively is 72%).

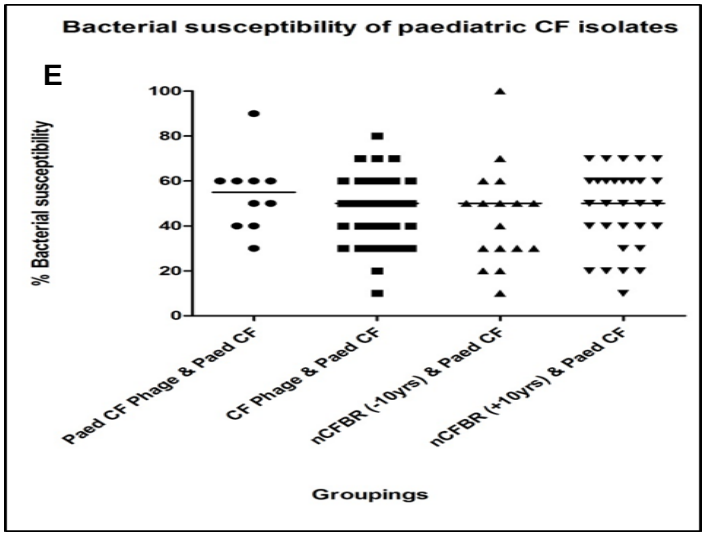
Even though the previous modelling and infection heat plots show the differences in infection profiles it does not show the distribution of these infectivity profiles e.g. the numbers of isolates within each aetiology the phages can infect. Figure 3.6 breaks down the data in figure 3.1 and stratifies it to the disease aetiologies by looking at the phages infectivity induced from these bacterial isolate figure 3.6 a-d and also the bacterial susceptibility to the phages figure 3.6 e-h. From figure 3.6 a-d it is evident that the CF phages have the highest infection range as a percentage. Figure 3.6.f shows that CF phages more readily infect their originating host when compared to BR phage (figure 3.6.h). The number of samples in each clinical disease groups varies and thus the data is not evenly distributed. However, the greatest variation of infection is seen in the BR isolated phages. The data in figure 3.6 a-d shows that the CF phages have the highest infection profiles when compared to paediatric CF and < 10 BR phages. This could be due to the poly microbial nature of the lung or due to the adult CF phages being better adapted to infecting their host.

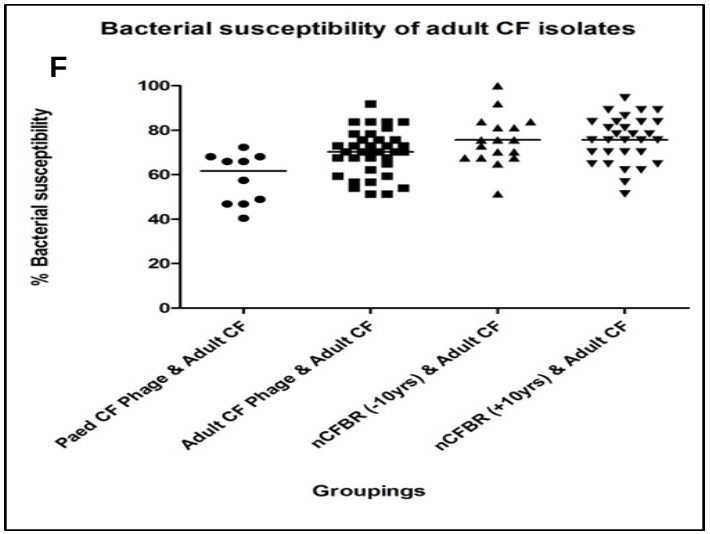

A
Phage infectivity of paediatric CF isolates
% Infectivity
100
80
60
40
20
0
Paed CF Phage & Paed CF
CF Phage & Paed CF
nCFBR Phage (-10yrs) & Paed CF
nCFBR Phage(+10yrs) & Paed CF
Groupings

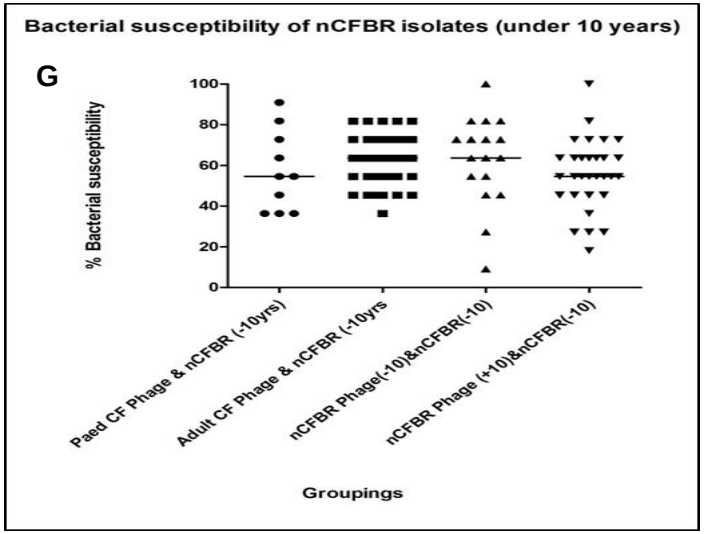

B
Phage infectivity of adult CF isolates
% Infectivity
100
80
60
40
20
0
Paed CF Phage & Adult CF
Adult CF Phage & Adult CF
nCFBR Phage(-10yrs) & Adult CF
nCFBR Phage(+10yrs) & Adult CF
Groupings


Phage infectivity of nCFBR isolates (under 10 years)
C
% Infectivity
100
80
60
40
20
0
Paed CF Phage & nCFBR (-10yrs)
Adult CF Phage & nCFBR(-10yrs)
nCFBR Phage(-10)&nCFBR(-10)
nCFBR Phage(+10)&nCFBR(-10)
Groupings


Phage infectivity of nCFBR isolates (over 10 years)
D
% Infectivity
100
80
60
40
20
0
Paed CF Phage & nCFBR (+10yrs)
Adult CF Phage &nCFBR (+10yrs)
nCFBR Phage(-10yrs)&nCFBR(+10)
nCFBR Phage(+10yrs)&nCFBR(+10)
Groupings


Bacterial susceptibility of paediatric CF isolates
E
% Bacterial susceptibility
100
80
60
40
20
0
Paed CF Phage & Paed CF
CF Phage & Paed CF
nCFBR (-10yrs) & Paed CF
nCFBR (+10yrs) & Paed CF
Groupings


Bacterial susceptibility of adult CF isolates
F
% Bacterial susceptibility
100
80
60
40
20
0
Paed CF Phage & Adult CF
Adult CF Phage & Adult CF
nCFBR (-10yrs) & Adult CF
nCFBR (+10yrs) & Adult CF
Groupings


Bacterial susceptibility of nCFBR isolates (under 10 years)
G
% Bacterial susceptibility
100
80
60
40
20
0
Paed CF Phage & nCFBR (-10yrs)
Adult CF Phage & nCFBR (-10yrs
nCFBR Phage(-10)&nCFBR(-10)
nCFBR Phage (+10)&nCFBR(-10)
Groupings

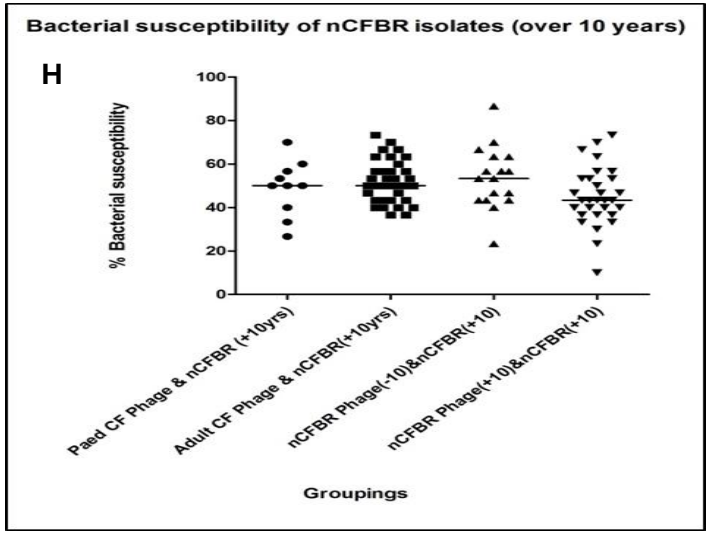

Bacterial susceptibility of nCFBR isolates (over 10 years)
H
% Bacterial susceptibility
100
80
60
40
20
0
Paed CF Phage & nCFBR (+10yrs)
Adult CF Phage & nCFBR(+10yrs)
nCFBR Phage(-10)&nCFBR(+10)
nCFBR Phage(+10)&nCFBR(+10)
Groupings

Figure 3.6: Graphical representation of bacterial sensitivity to infection and phage infectivity based on disease aetiologies (BR is referred to as nCFBR). A graphical representation of the results in Figure 3.1. A) The infectivity of phage lysates on paediatric CF hosts. B) The effect of phage lysates on adult CF hosts. C) The effect of phage lysates on BR hosts (less than 10-year *P. aeruginosa* infection). D) The effect of phage lysates on BR hosts (over 10-year *P. aeruginosa* infection). E) The susceptibility of paediatric CF bacteria to phage lysates. F) The susceptibility of adult CF bacteria to phage lysates. G) The susceptibility of BR bacteria (under 10-year *P. aeruginosa* infection) to phage lysates. H) The susceptibility of BR bacteria (over 10-year *P. aeruginosa* infection) to phage lysates.

3.3 Discussion

3.3.1 Cross Infection of Pa temperate bacteriophage/phage

This study is one of the largest cross infection profiling of temperate bacteriophage and their hosts reported. In this study every strain of Pa isolated from the lung environment was shown to harbour at least one inducible temperate phage as they were able to infect at least one other bacterial isolate. The approach of using potentially mixed phage lysates to generate a large-scale infection profile for each lysate and total host range across the 94 Pa clinical isolates is also a novel approach. Previous studies looking at phage host range tends to plaque purify the phage by at least 2 rounds of picking individual plaques to propagate. This is not a representation of what happens in the chronic lung as the stimuli for phage induction is global as built around the stress response of the bacterium. We hypothesise therefore that this offers a truer representation of what would happen in lung on induction of the bacterial SOS response. Temperate bacteriophages from the same bacterial host exhibiting varying infection profiles has been reviewed (Weitz et al, 2013). However, as these phage lysates reported here are not purified phages, they mimic the reservoir of viruses within the bacterial cell that can release and disseminate in the lung which is innovative.

The data in the cross infection study was extrapolated and presented as a heat map derived in R (Figure 3.1). The heat map aids visualisation of the raw infection profile in appendix 1. The same data was shown using the nestedness and connectance plot in Figure 3.2a. Both of these figures show more conveniently that the CF derived phages are more readily able to infect than their BR derived counter parts indicated by having a greater nestedness. The data was not stratified into the disease aetiologies as no trend was observed. Figure 3.2 characterises the interactions between the phage lysates and their Pa bacterial host. It can be inferred therefore that the change in connectance implies that the infection difference is driving the variance. The increased nestedness shown by the CF phages supports our hypothesis that the

adult CF phages either individual or per lysate have evolved a broader host range and higher infection rate across the whole Pa panel. Other clinical data was modelled to see if there were significant differences between the phage infection profiles although nothing except mucoid phenotype was identified as significant using both bi-partite modelling and PLS-DA.

Figure 3.2b shows that the BR mucoid (n=22) derived phages could infect more readily, Pa that display a non-mucoid phenotype. This suggests that the phages may have evolved infection strategies to infect this phenotype of Pa isolate. Phages have been shown to be blocked from infection by bacteria producing a mucoid phenotype. Examples of this have been seen in *Salmonella anatum* and phage restriction by mucoid phenotype (Mcconnell et al, 1986). It has also been illustrated that Pa temperate phages have previously been identified and characterised that can infect mucoid isolates, e.g. Ab31 a chimeric phage that is related to other temperate Pa phage PAJU2 and *P. putida* phage AF (Latino et al, 2014).

The clinical data available was used to compare between these phage lysates and sensitivity to phage infection (Tables 2.2.1 and 2.2.2) and modelled the data based on the clinical variables. It was determined that the data supports our cross infection study and further supports our theory that phage and their host evolve together in the lung of BR and CF patients, where infectivity of phage and susceptibility of the host can be used as a marker of the aetiology of disease, mainly age (CF) and time since incidence (BR). The Norman et al., (2016) paper shows that an increase in presence of *Caudovirales* can be a marker of gut disease (Norman et al, 2015).

PLS-DA here has been used to elucidate the difference in phage infection profiles compared to the aetiology stratifications of the disease states. The data illustrates that phage from BR Pa isolates have a different infection profiles when compared to phage from CF Pa isolates (Figure 3.1, Figure 3.2 and Figure 3.3). One driving factor could be the site of temperate phage adsorption (Chibeu et al, 2009). Pa phages have been shown to adsorb to the Pa twitching pili (Pemberton, 1973). It has

been previously shown that in the CF lung, Pa that chronically infect the lung lose the ability to twitch over time as the trait is not needed (Comolli et al, 1999). It may be possible that adaptation to use an alternate receptor is a more efficient way of infecting the cell within the lung. Another factor could be due to phage resistant mechanisms (see chapter 1.6.6 for more detail of phage resistant systems). An example of this is CRISPR/CAS systems that the host Pa use to recognise and abort phage lysogenic infections (Villion & Moineau, 2013). Phages evolve to develop anti CRISPR/CAS systems to evade host detection and are able to super infect the host (Bondy-Denomy et al, 2013). Thus it is hypothesised that BR phage may not be as evolved when comparing CF and their phage evolution and hence why lower infectious rate is seen of BR phage re-infecting their originating hosts as have not adapted or evolved this mechanism. It may also be that the pathophysiology of the lung is such that there is difference in the adaptation and evolution in each disease. There is also a noticeable difference between CF phage from paediatric and adults (Figure 3.1, 3.2 and 3.3). The data suggest that CF paediatric phage and BR phage from less than 10 years have a similar profile. The different models Figures 3.3, 3.4 and 3.5 stratify BR phage from CF phage supporting our hypothesis that they are from different disease state and one prediction is that the linking factor is possibly due to prolonged evolutionary arms race between the host and the phage. It could also be hypothesised that the bacteria are as evolved based on their current disease states, phages evolving alongside their host may become to specialised to the host and narrow their host range. A study artificially looking at coevolved phage bacteriophage ϕ2 infecting *Pseudomonas fluorescens* SBW25 showed as one of their finding that the host range become more resistant to genetically similar phage it co-evolved with (Gorter et al, 2015). This coevolution based host resistance is not indicative of the data presented in this thesis.

Interestingly a higher number of CF Pa phages when spotted onto their originating hosts could re-infect compared to the BR phages (figure 3.6.e-h). This finding goes against the global convention of temperate phage infection; the prophage

repressors function is to prevent super infection (Kholodii & Mindlin, 1985). It was previously thought that if a phage was isogenic such that it was very similar at a genotypic and phenotypic level then it would not be possible to infect the same host again (Kaiser & Jacob, 1957). This suggests that the CF Pa phages and their hosts have diversified together and hypothetically, they are more evolved to re-infect their hosts in this environment. Possibly due to the lower lung environment being constrained and thus re-infection is supportive of positive selection for the phage. We could hypothesise that multiple isogenic infection leads to gene addition and increased expression of genes that increase the selective advantage. Fogg et al. (2011) illustrated that multiple copies of Stx-phage 24B increased the rates of lysogeny with each successive infection and also that gene addition of the phage could increase levels of the Shigatoxin (Fogg et al, 2011). These phages mediated traits may offer selection with exacerbation of symptoms as a result of the patient's inflammatory responses causing tissue scarring coupled with poor movement of the mucus and the ability of Pa to readily form biofilms. In *E. coli* it was shown that the phage 24B could re-infect its host multiple times with increasing frequency on each round of infection (Allison et al, 2003). A bacterial host harbouring multiple phage could increase competitiveness (Burns et al, 2015) and in the case of *Pseudomonas* this is well described with certain phages increasing infection whilst other preventing infection, see chapter 1.4 for more detail (Winstanley et al, 2009).

In Figure 3.6 panels A-D, illustrates that the adult CF phage can infect all the other host bacteria with nearly 100% infectivity. However, the infection rate back onto their host is lower. The paediatric phage also appears to have the lowest median percentage infectivity, this may indicate that these phages have not evolved in these environments and so are still recognised by bacterial defence systems, indicating phage naivety. This pattern of naivety is opposite in the BR phage infections, the phage from BR (under 10 years) have a higher median percentage infectivity compared to the phage from BR (over 10 years), it would appear that the phages lose

their infectivity over time but this could also correlate with total patient deterioration and poor health, as the onset of BR is indicative of an older population. Or it may be that there is a clonal population in the lung, previously reported in the CF lung (Haussler et al, 2003) that means diversity in infection profile is not needed. Panels E – H in Figure 3.6 demonstrate the susceptibility of the host bacteria to phage infection.

This study illustrates another level of complexity in phages from these disease states where despite phages and their Pa host have genome similarities and narrow host range, in terms of temperate phages, they still exhibit differences in their infection profiles and phage infection susceptibilities (Ceyssens et al, 2011). This can be hypothetically linked to the selective pressure the temperate phage and their hosts are under within the CF lung.

4. Isolating single *Pseudomonas aeruginosa* phage genomes from metagenomic DNA sequencing.

4.1 Introduction

Identification and characterisation of bacteriophages is continuing to show the diversity of bacterial viruses that are present in the biosphere. A wide range of methods have been used to characterise and classify phages. Bacteriophage morphologies have been widely described (figure 1.2a and figure 1.2b). When working with bacteriophages it is important to characterise the genome and marry genotype to phenotype and biology. Phage genomics can also be important to classify downstream analysis such as profiling for potential alteration through mutation or recombination. However, the major concern regarding temperate phages is their influence on bacterial evolutionary selection and how this correlates to the physiology of the bacteria and how this contributes pathogenicity (chapter 1.6.7.1). Some studies have shown that increased virulence of a bacterium can be as a result of bacteriophage carriage (also discussed in the introduction of chapter 3.1). An example is Shiga toxin encoding *E.coli* (STEC) 0157:7 is a strain that encode the Stx 1/ Stx 2 that is encoded by Stx phage (Allison, 2007; Plunkett et al, 1999). A filamentous phage, CTXϕ is shown to be responsible for coding the cholera toxin in *V. cholerae* (Waldor & Mekalanos, 1996). Similarly some bacteriophages can be responsible for antibiotic resistance genes via transduction (Abedon, 2011), although identification and function of these genes have been brought into question. Enault et al. (2016) describes how phage encoded genes transferring true antimicrobial resistance is lower than what was previously thought (Enault et al, 2016).

PAO1 has a genome of 6.3 Mbp (Stover et al, 2000) and Liverpool epidemic strain LESB58 has a genome size of 6.6 Mbp (Winstanley et al, 2009). One reason for the genome expansion is LESB58 strain carrying 5 genomic islands and 6 prophage regions (Winstanley et al, 2009). Further research on the inducible phages and their biology has been well reported and discussed in chapter 1.4.

When characterising the genome of temperate phages it can be important to identify genes associated with lysogeny. This includes but is not limited to the *Cro, CI, CII* and integrase genes. It can be important to establish using genome characterisation that the phage resembles, at the genome level, similarities to other known temperate phage genes. A study by Braid et al. (2004) compared a Pa phage B3 with D3112 and Mu as it shared 7.5 kb homology in the right arm of the chromosome. Importantly they found genetic rearrangement in the left arm of the chromosome (Braid et al, 2004). It can be therefore important to look at the genome architecture, which can be used to potentially identify recent recombination events that may be evident based on comparison to taxonomically similar phages. Smith et al. (2012) describes the conserved architecture and genome organisation of dsDNA lambdoid-like phages. This can be used as an important tool to identify similarities and differences between phages. Importantly this organisation of genes is invaluable for genome assembly, especially in metaviral populations.

The predominant phages described to infect *Pseudomonas aeruginosa* (Pa) are part of the *Cauldovirales*, double stranded, tailed, DNA bacteriophages/phages. Previously phages have been categorised and characterised based on their morphology (chapter 1.6). Taking advantage of sequencing technologies, it can be more useful to compare phage genomes to identify alteration and diversification within genomes. Comparative genomics of phage genomes enables one to identify genetic diversity between phages. The genome architecture can be compared, which is mostly mosaic suggesting high degree of horizontal gene transfer (Hatfull, 2008). Studying phage genomics is an area of novel genetics as genes are of mostly small with unknown function at both the nucleotide and protein level (Hatfull, 2008; Smith et al, 2012).

Stratifying metaviral DNA sequence data allowing assembly of individual viruses is classed as difficult due to mosaicism and genes with similar sequence or function that could offer chimeric phage on assembly. Genome assembly is also difficult due to lack of reference genomes submitted to databases. A study looking at four exhaustive

iterative assembly, CLC Genomics Workbench 6.0.4 assembler, Genovo version 0.4, and Newbler 2.5 identified that all the exhaustive iterative assemblers were capable at assembling contigs of viral metagenome (Smits et al, 2014). However, the study went on to suggest that it was best to use a combination of assemblers to complete assembly of genomes (Smits et al, 2014). Thus iterative assembler like PRICE, IDBA-UD, Velvet and SPAdes are seen to be best at assembling viral metagenomes (Smits et al, 2015).

The aim of this chapter was to develop a protocol to isolate individual phage genomes from a mixed virus sample. We aimed to test the hypothesis that using the biology of the bacteriophage including burst size or preferential integration sites of lysogeny could be used to aid stratification of the DNA sequencing data that would allow individual assembly of phage genomes from a mixed viral community sample induced from a clonal bacterial population. The study would aim to add phage genomes to nucleotide database, which would potentially offer novel Pa phage genomes.

4.2. Results

4.2.1 Genome characterisation

In order to characterise the viral genomes, the phages were induced using the method described in 2.3.1. This method is a metagenomic approach to isolate all viral DNA that has been encapsulated as early steps removes extrachromosomal bacterial DNA through multiple rounds of DNase treatment. The presence of phage in the lysates was evaluated using the spot assay method described in 2.4.1 which is a quick way to assess phage infectivity on a known bacterial host.

Once phage presence was confirmed the samples were used in the Phage DNA isolation step described in 2.6.2. This step facilitates DNA extraction from the lysate as a meta phage isolation approach. A subset of the samples were tested for 16S rRNA PCR amplification. A low level of bacterial chromosomal contamination was determined by PCR for the 16S rRNA gene (Muyzer et al., 1993). Therefore in order to stop the carryover of bacterial genomic DNA a wet and bioinformatics approach was utilised. The first was the use a round of DNAse treatment described in methods chapter 2.6.2. The second was removing the error peak from the K-mer abundance. This was confirmed by assembling the error peak and on assembly and similarity search using local alignment (blastn) this error peak contains the remaining bacterial carryover.

The Pa clinical Isolates were renumbered (table 2.6.1) for convenience and to further blind the study.

We work under the broad hypothesis that phages integrate into preferential sites in the bacterial chromosome. They will also induce out in different burst sizes and we can use these two things as they link to sequencing depth to separate the k-mers based on abundance and count, separating the mixed phage genome sequences into their peaks.

Figure 4.1 shows an example of three samples, where 1, 3 and potentially 5 peaks were seen. All the k-mer graphs for the 94 Pa samples are presented in table 4.1.

In figure 4.1 the far left peaks are the k-mers associated with error and potentially bacterial chromosome. Distribution of k-mers, grouping together based on k-mer abundance and count, form peaks seen in figure 4.1. The k-mers associated with the peak were extracted using the “calculate median distribution” script described in the methods (2.6.5).

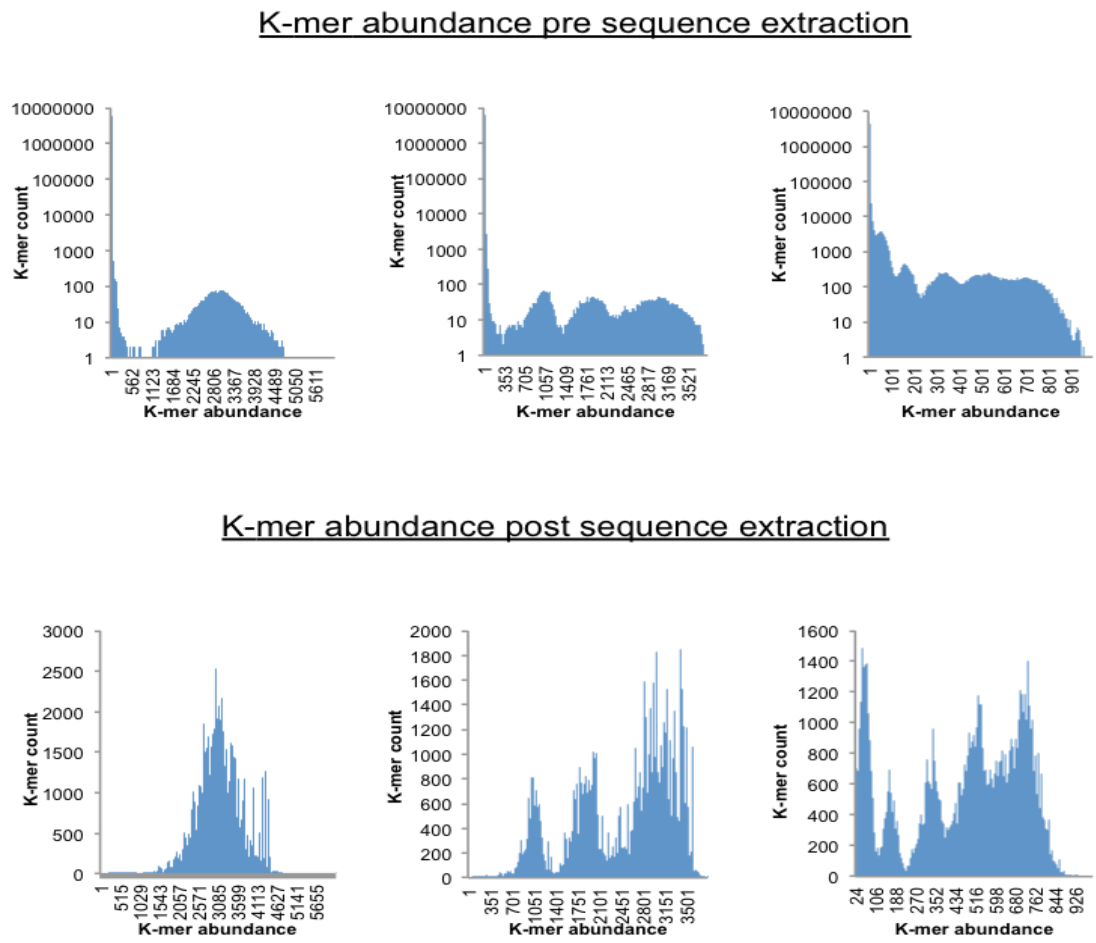
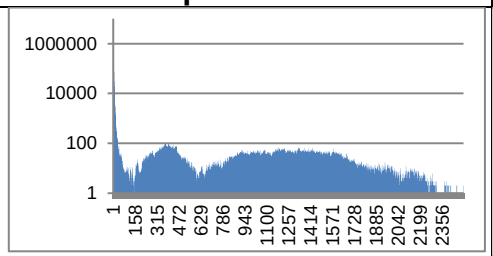
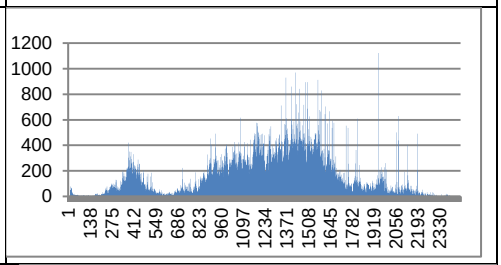
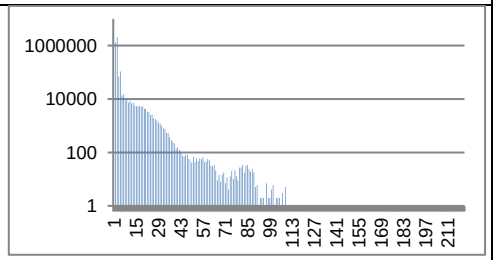
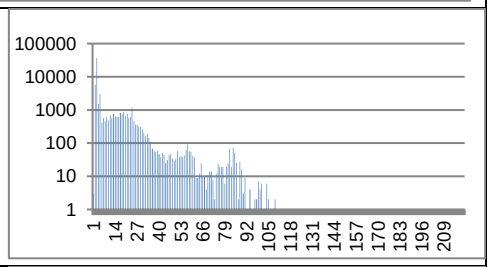
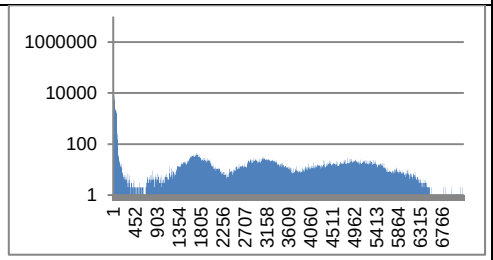
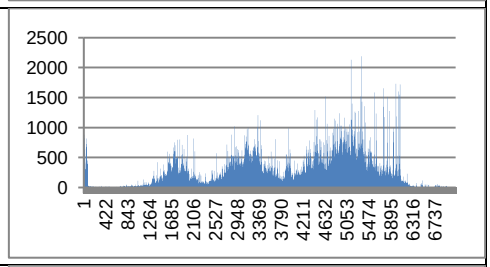
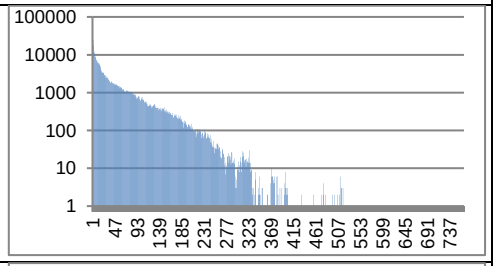
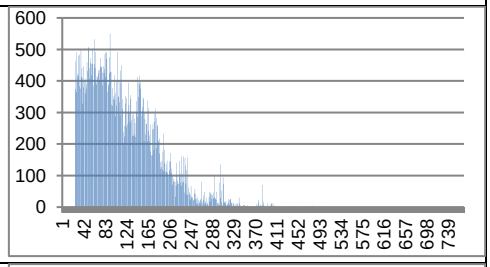
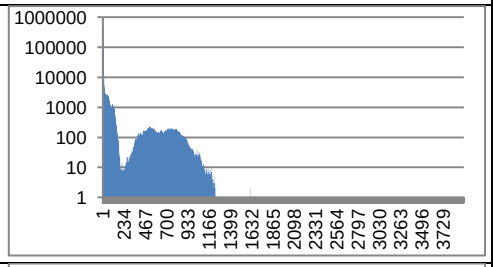
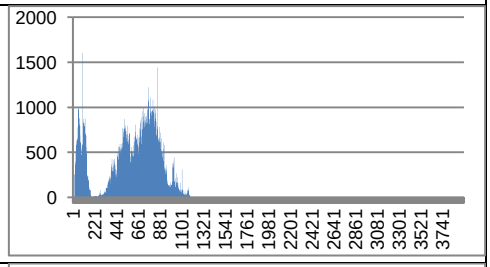
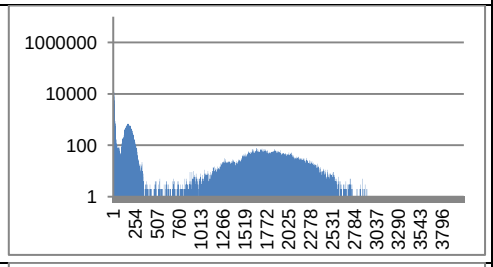
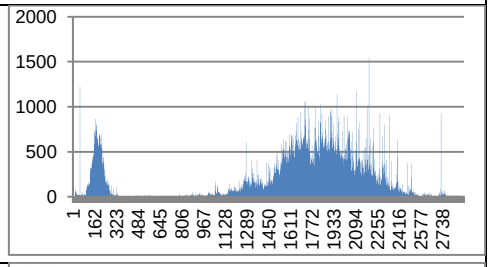
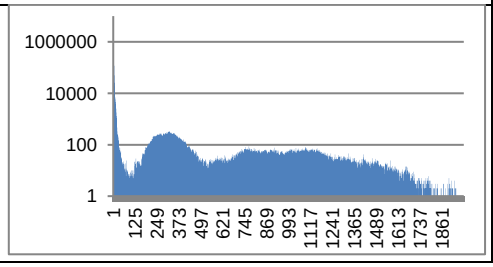
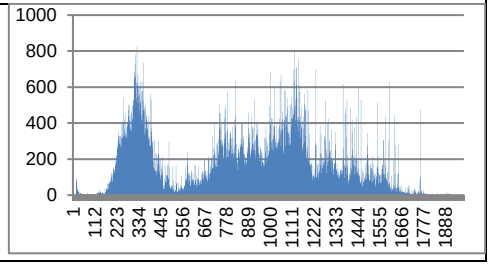
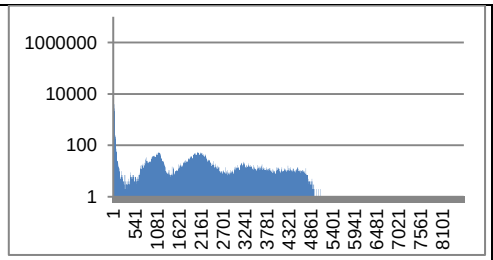
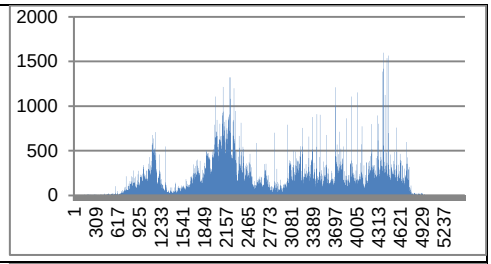
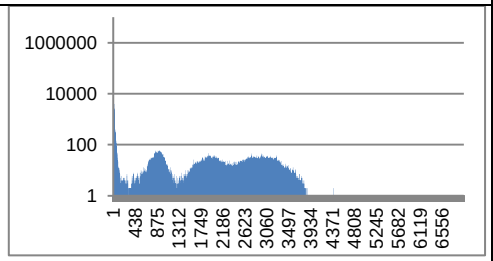
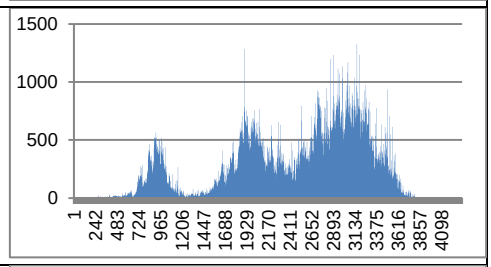
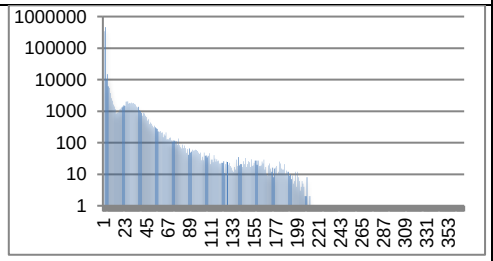
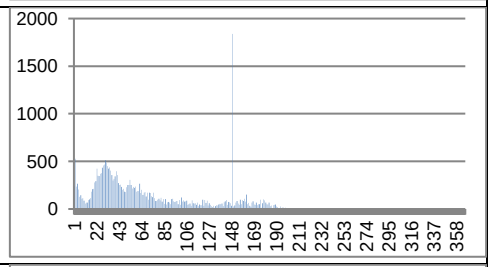
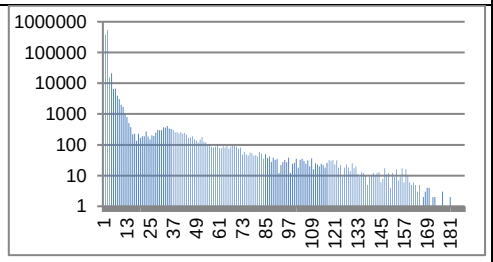
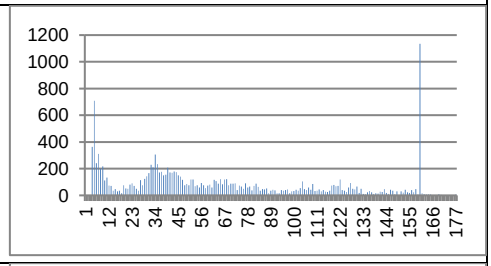
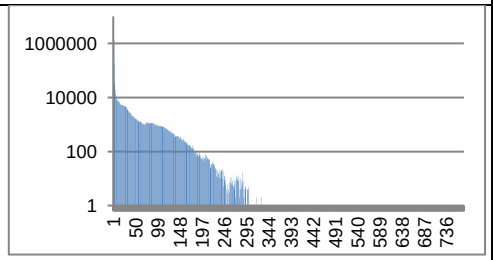
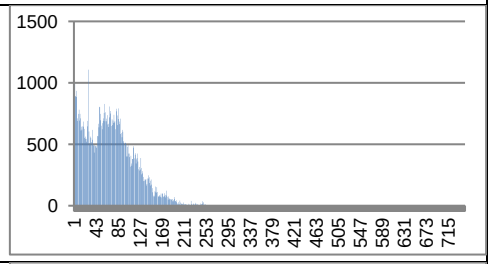
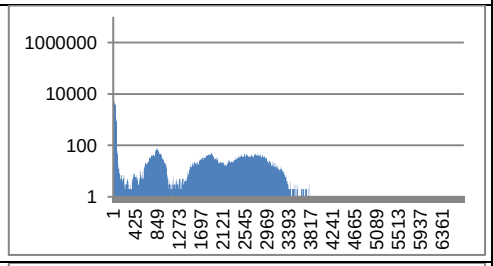
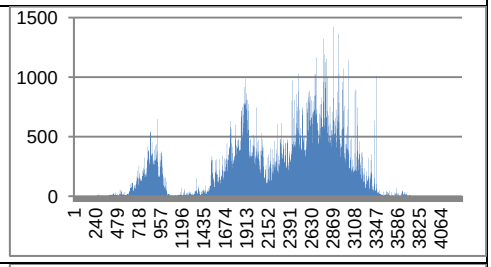
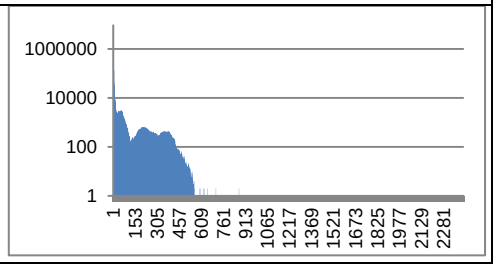
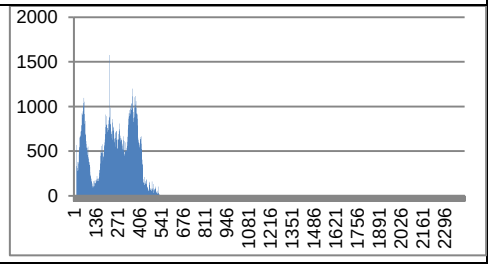
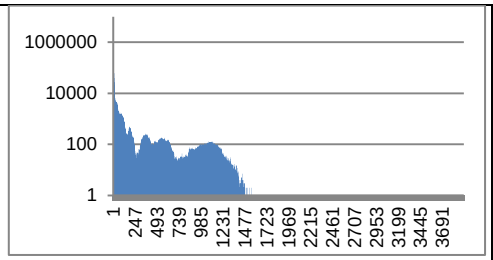
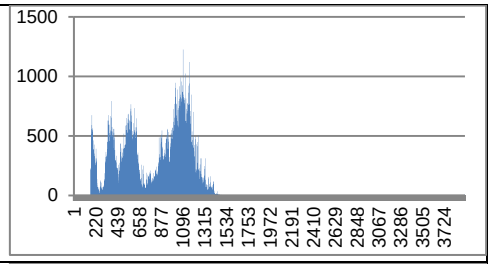
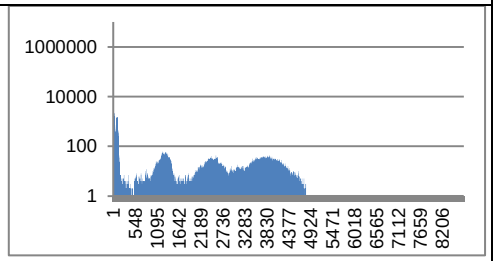
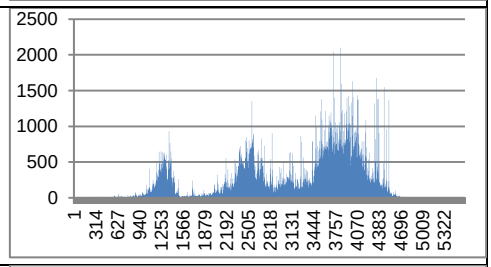
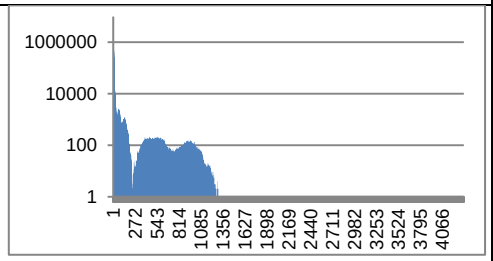
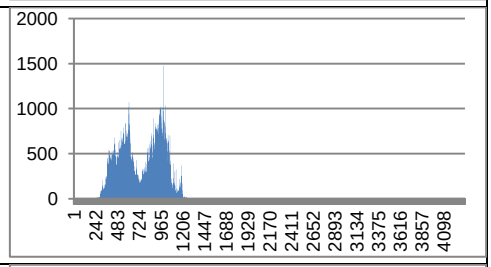
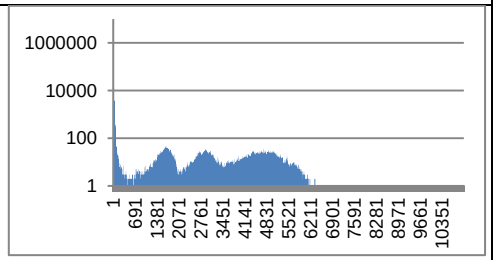
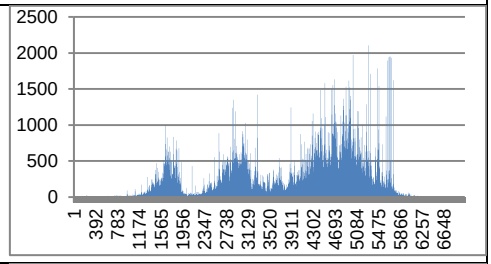
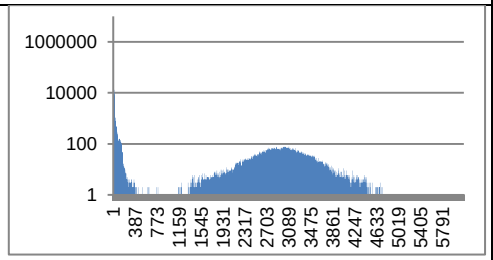
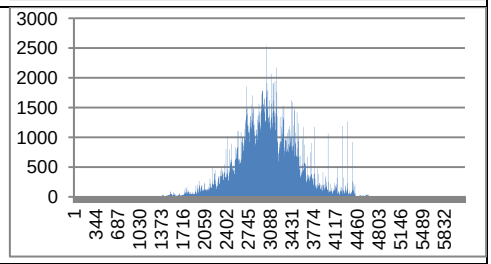
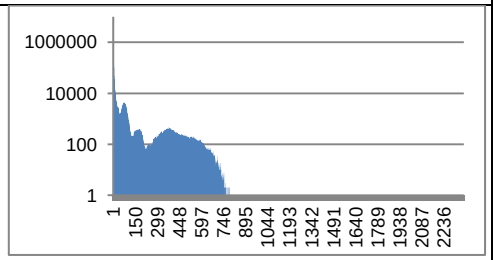
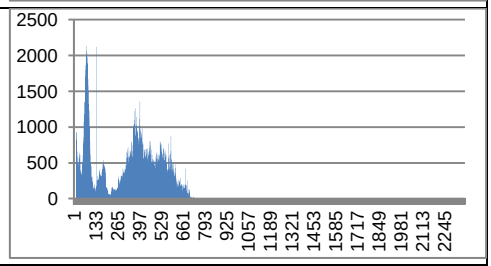
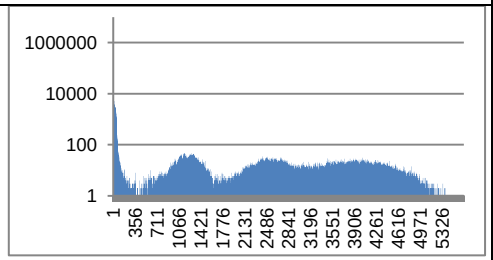
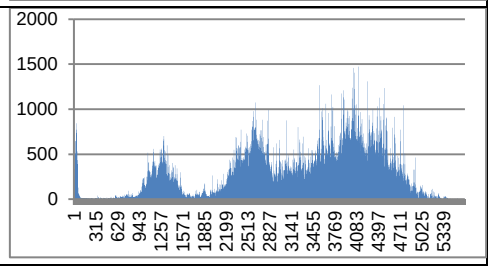
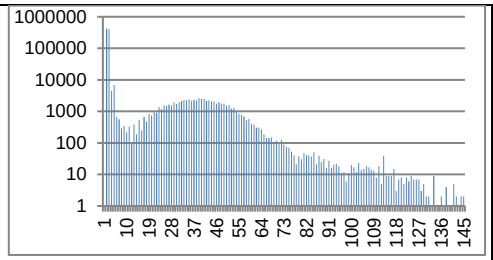
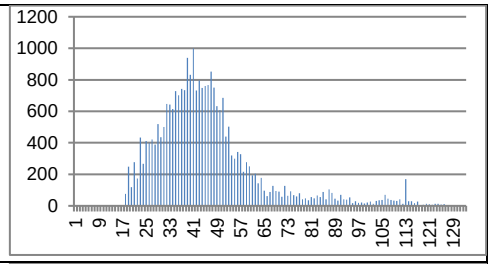
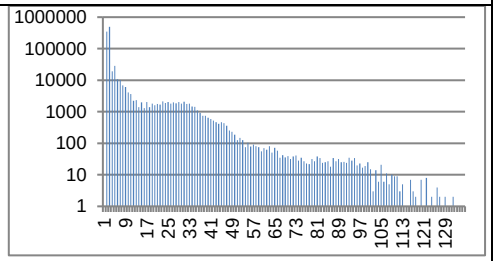
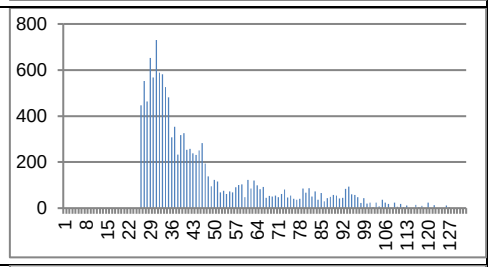
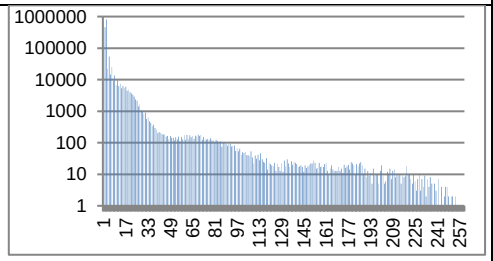
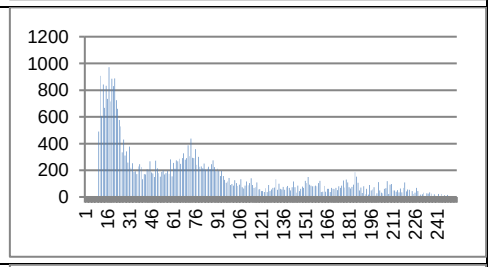
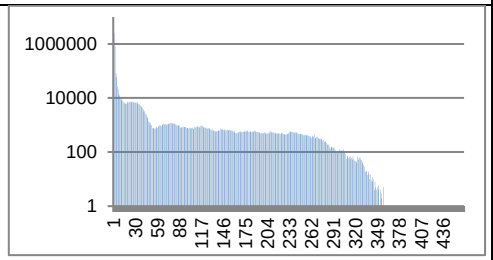
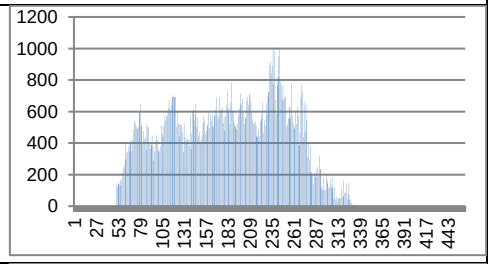
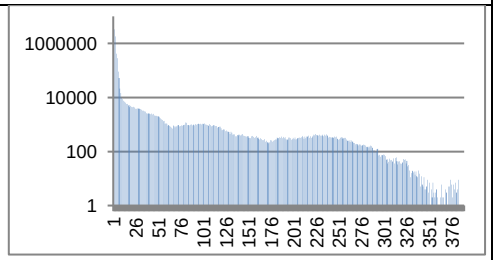
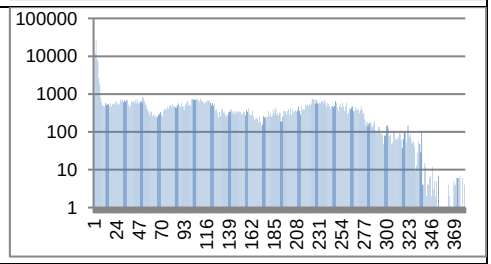
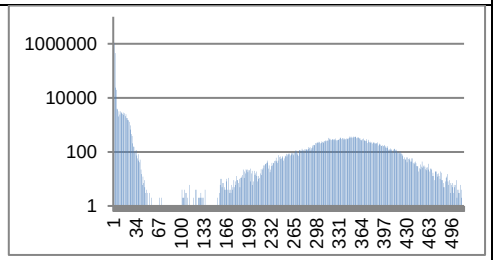
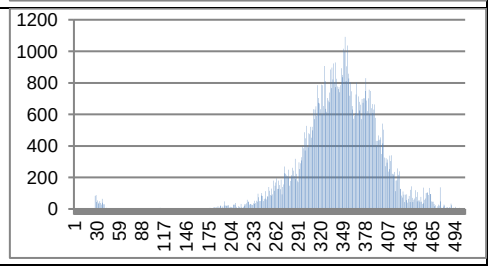
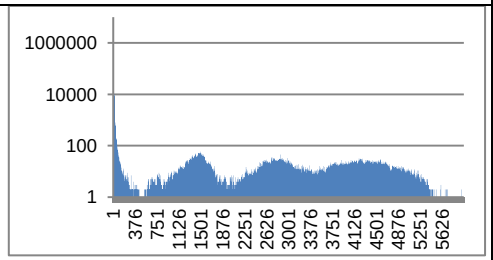
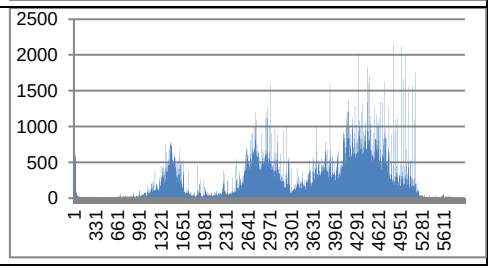


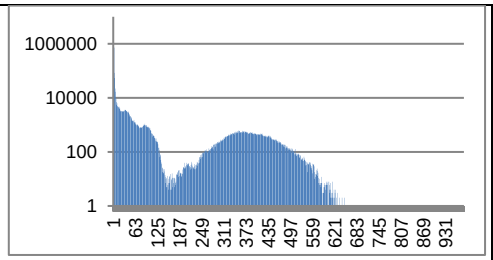
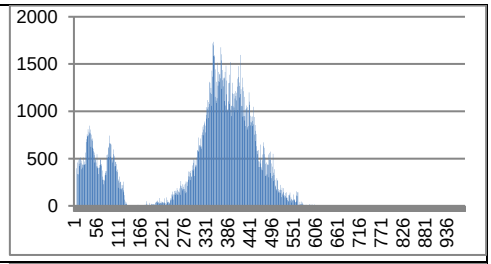
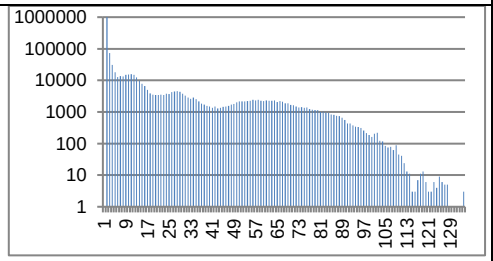
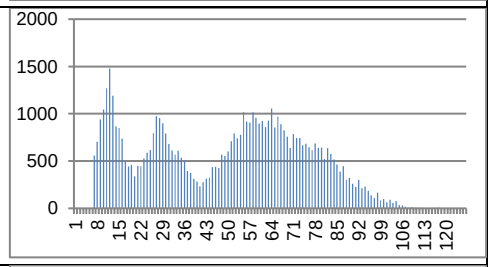
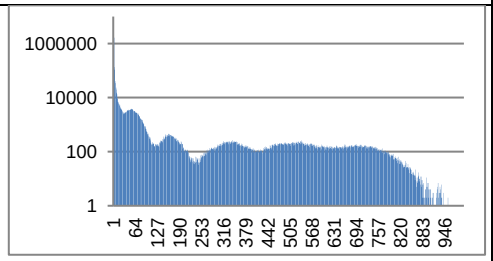
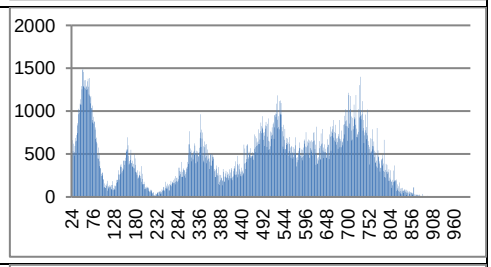
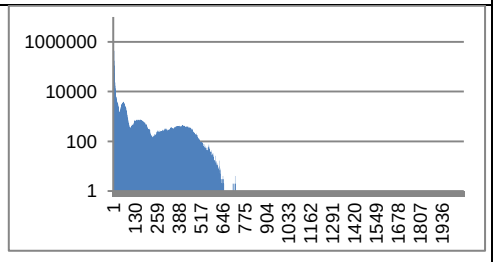
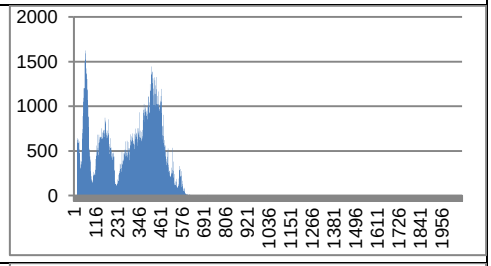
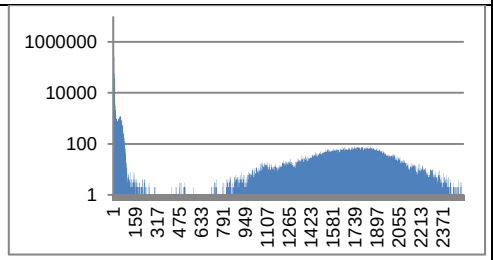
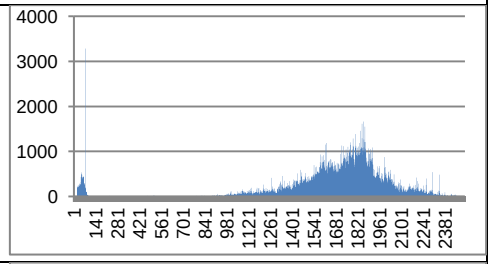
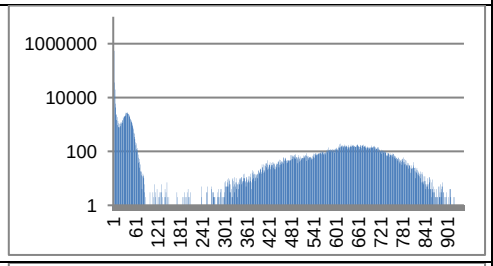
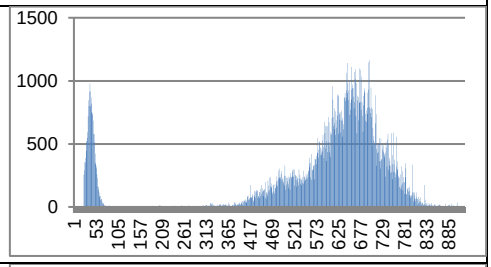
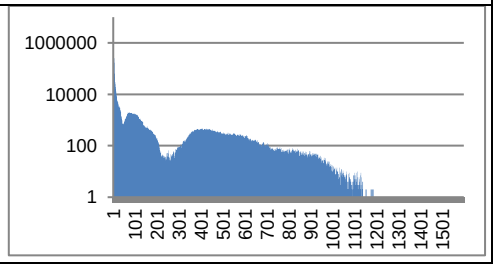
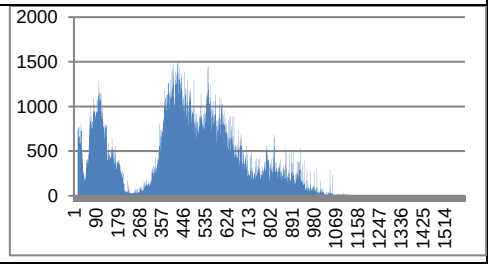
Figure 4.1: **K-mer abundance graph pre and post sequence extraction.** This figure exemplifies peak separation step, concurrently reducing potential bacterial and error related k-mer. The figure shows three examples of 1, 3 and potentially 5 peaks respectively.

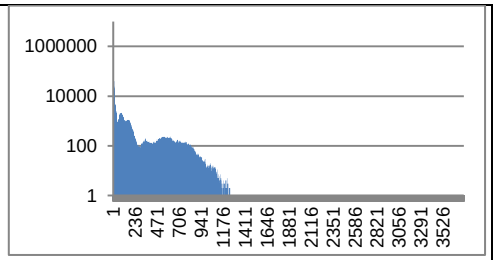
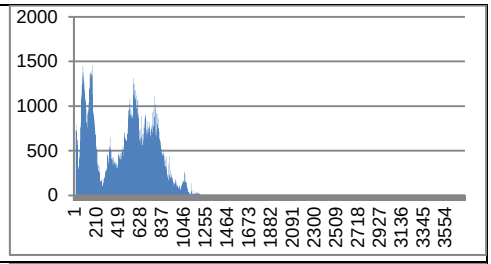
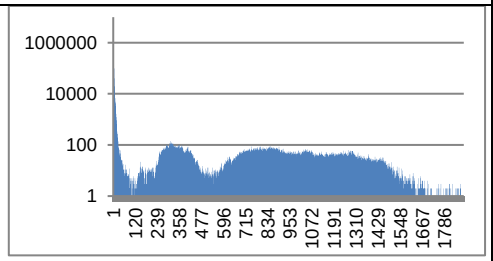
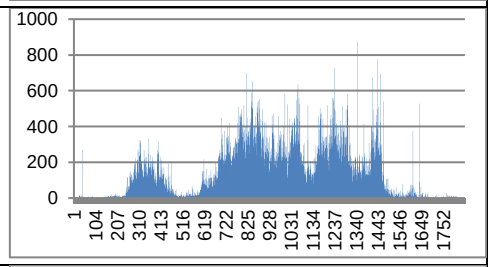
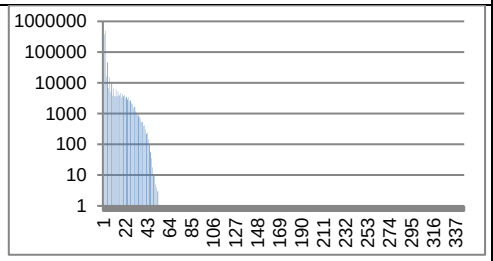
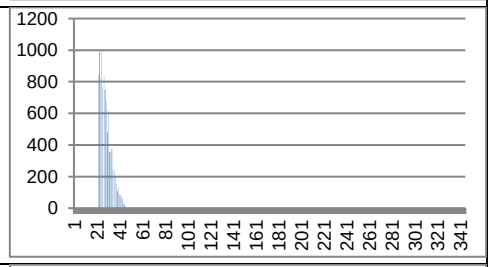
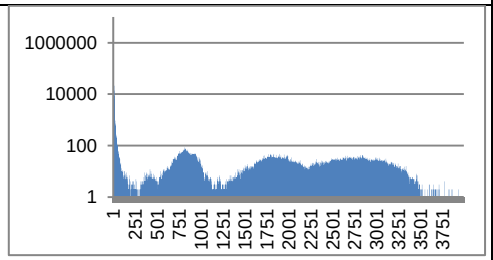
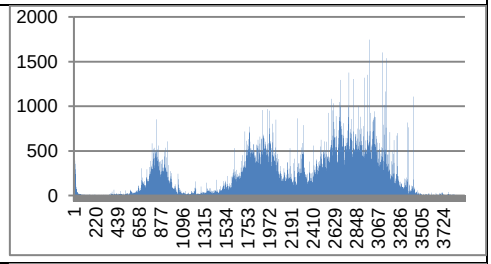
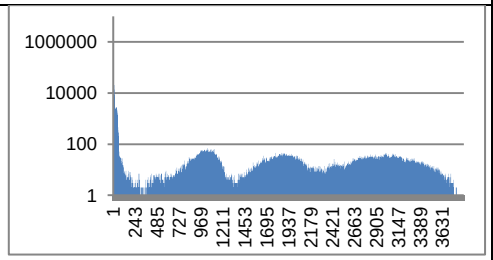
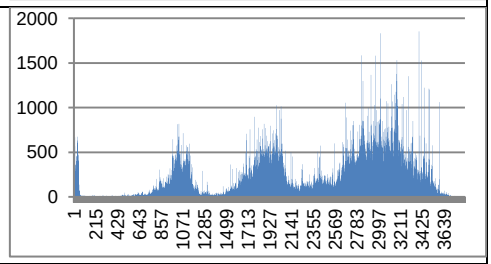
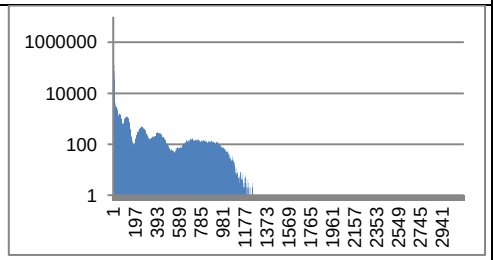
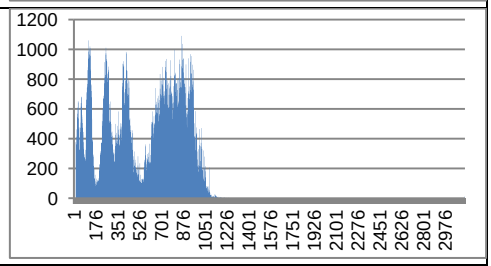
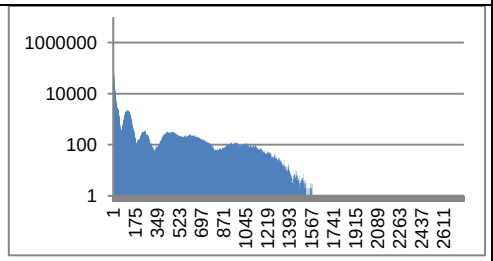
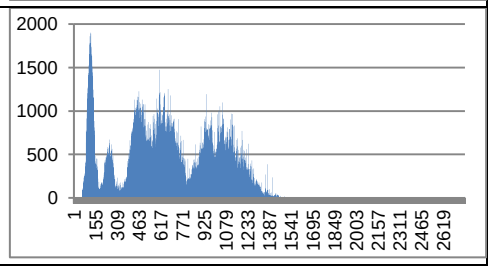
Sample number	Log scale of k-mer graph pre error peak removal	k-mer graph post error peak removal
1		
2		
3		
4		
5		
6		
7		

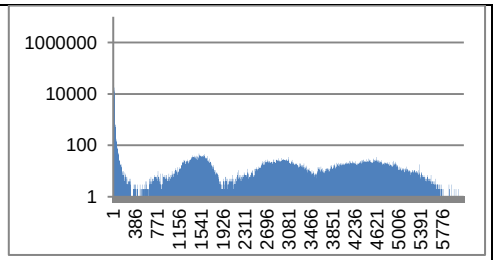
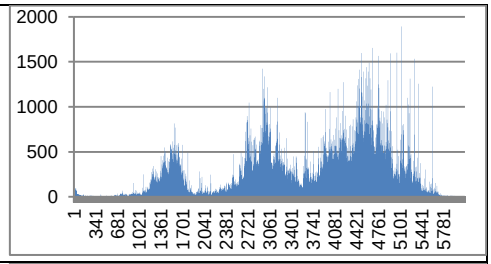
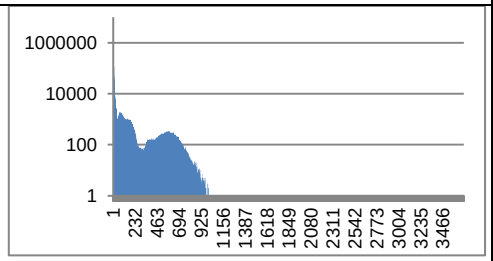
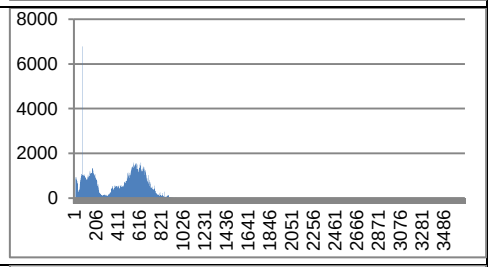
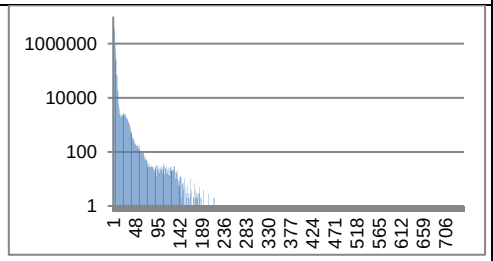
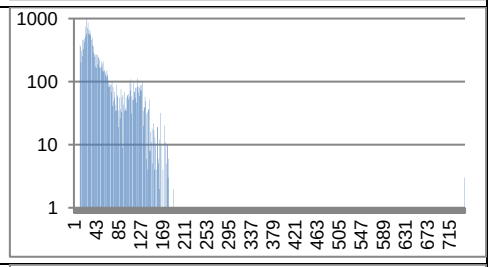
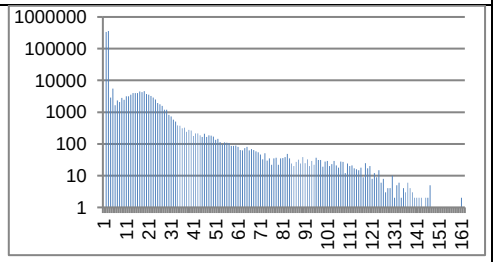
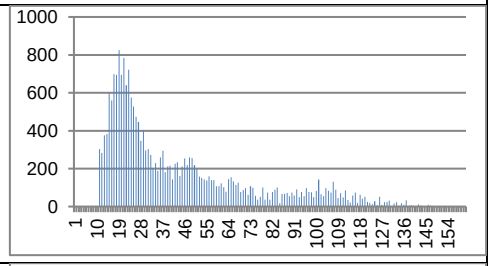
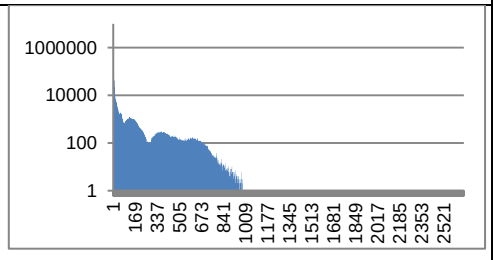
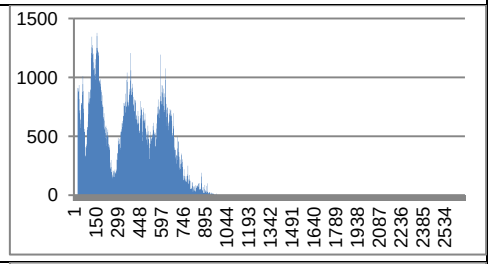
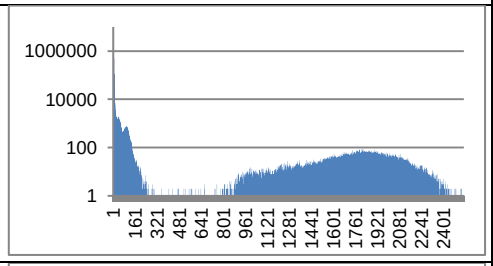
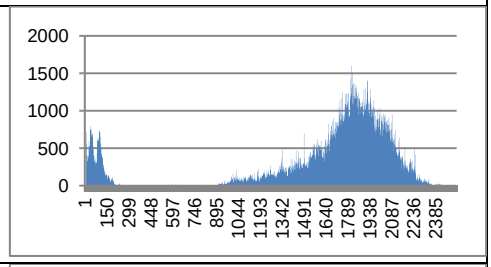
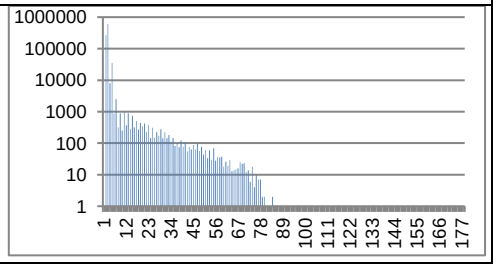
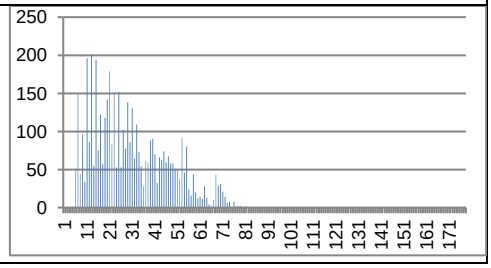
8		
9		
10		
11		
12		
13		
14		

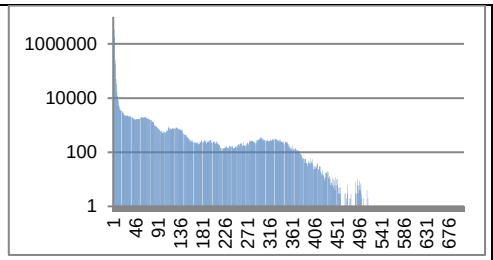
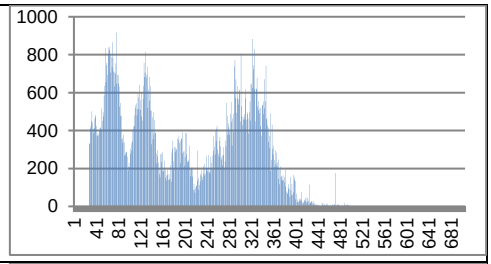
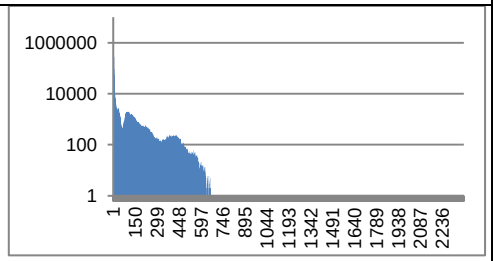
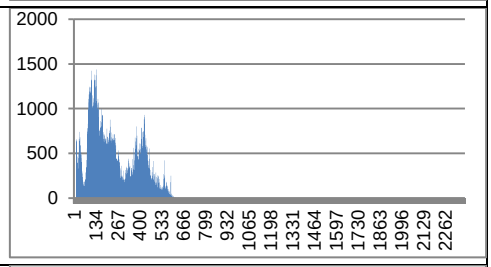
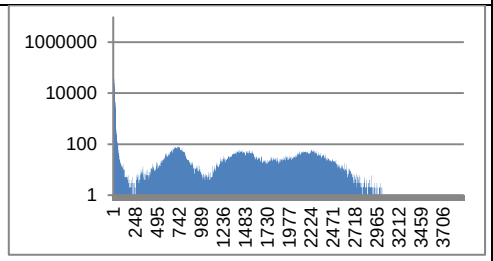
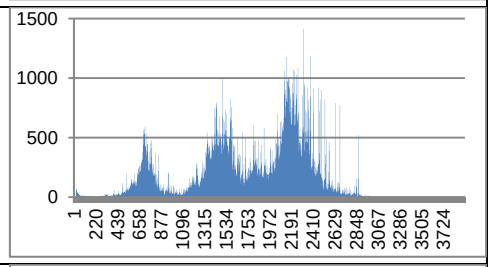
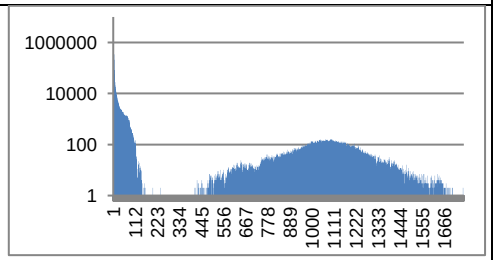
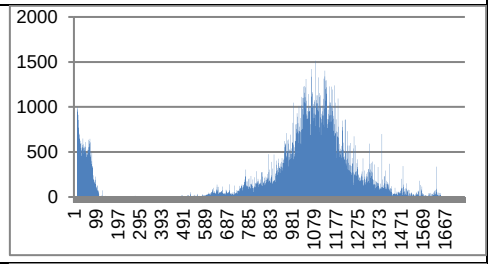
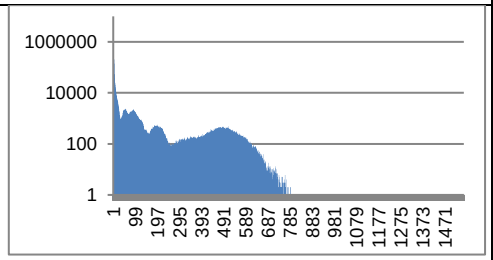
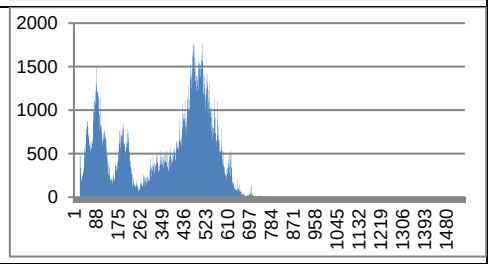
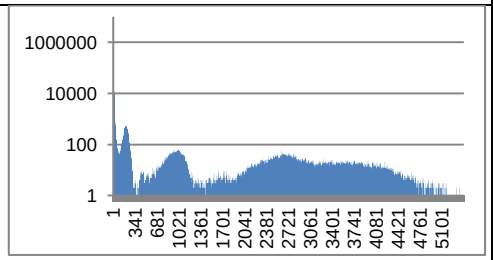
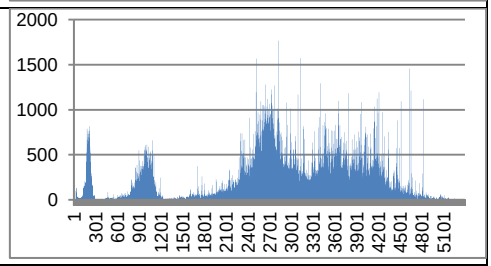
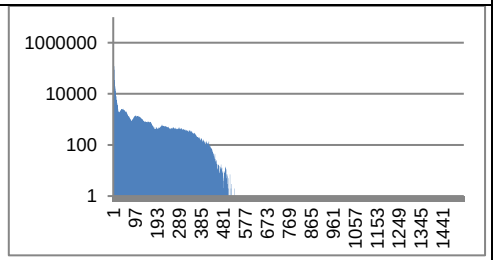
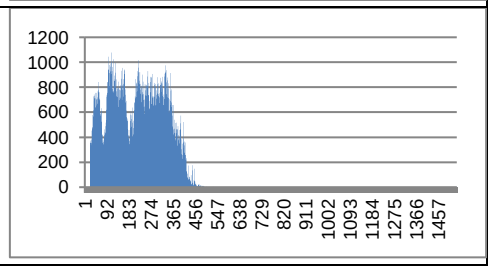
15		
16		
17		
18		
19		
20		
21		

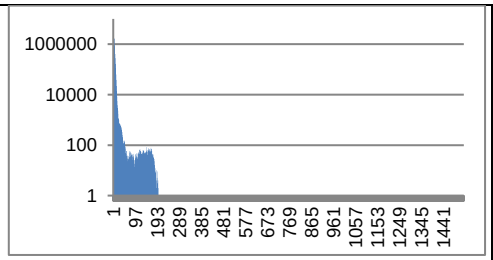
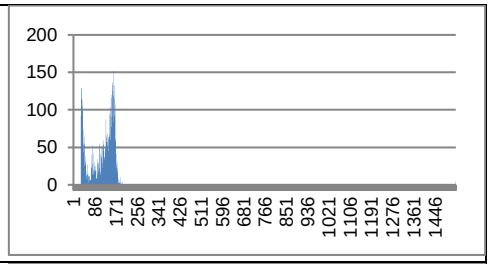
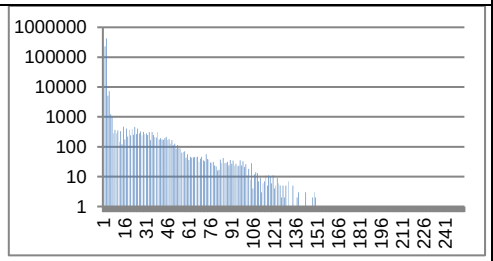
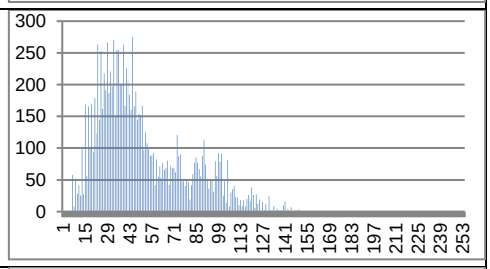
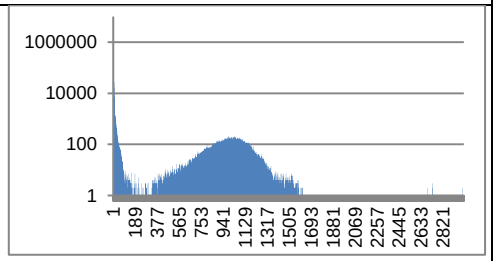
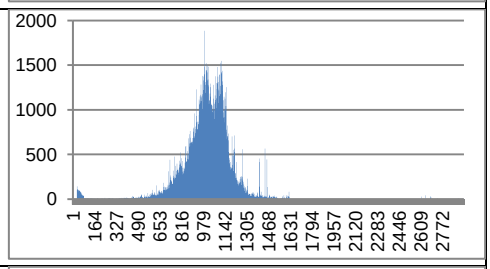
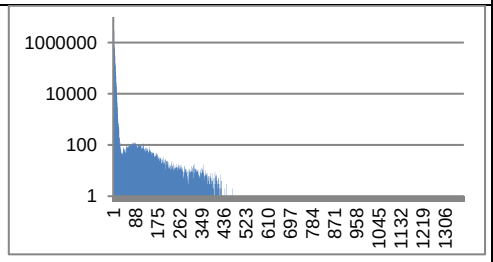
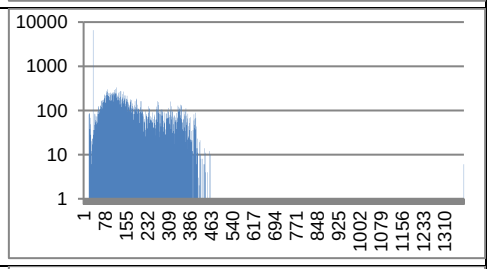
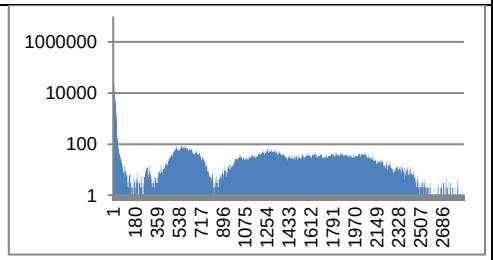
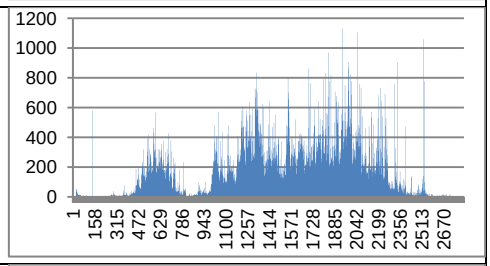
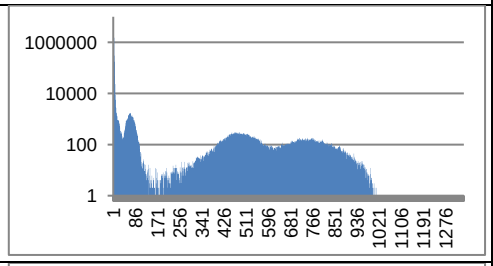
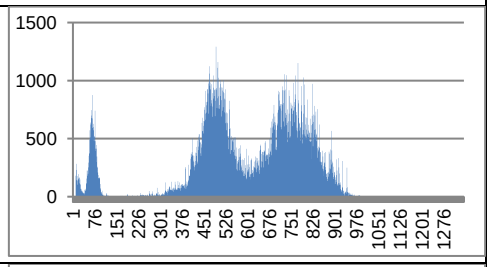
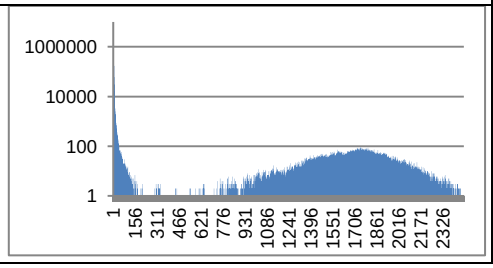
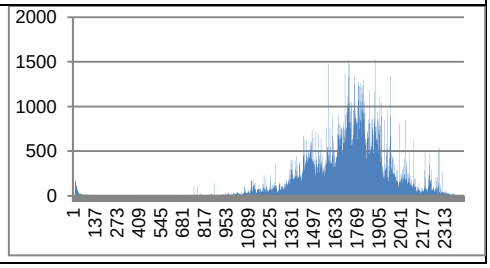
22		
23		
24		
25		
26		
27		
28		

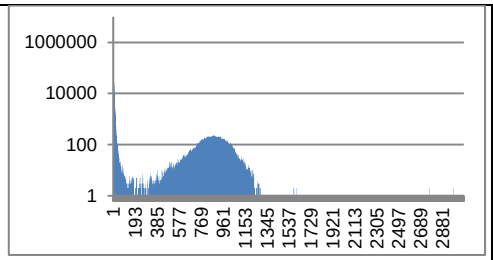
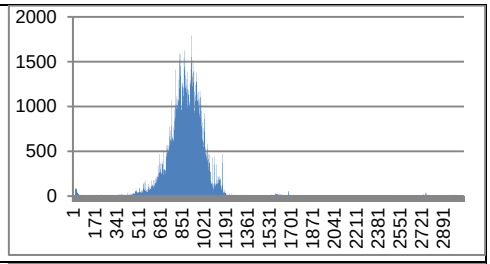
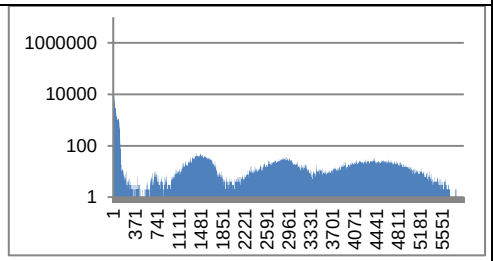
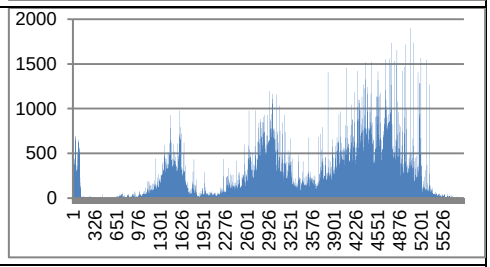
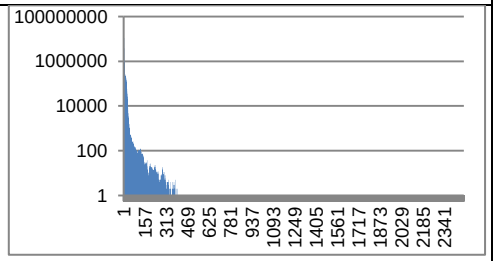
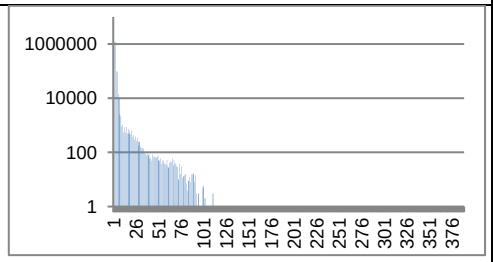
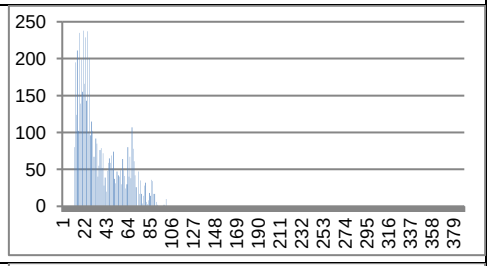
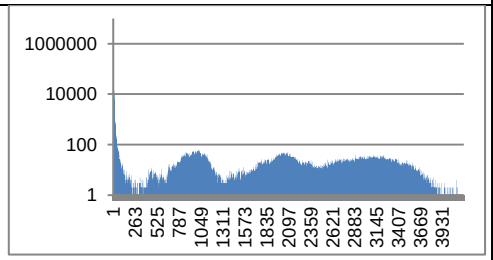
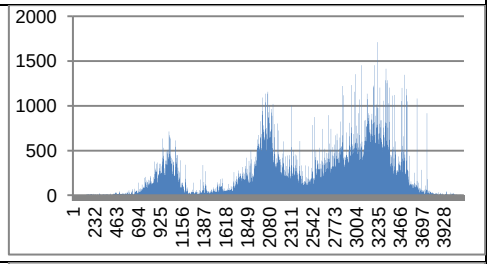
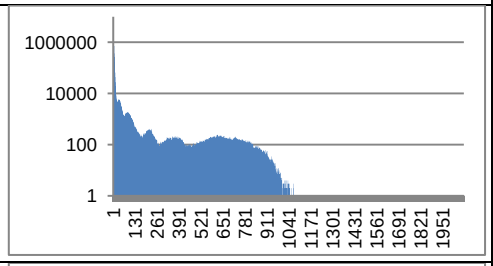
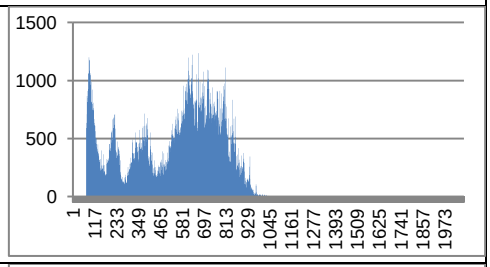
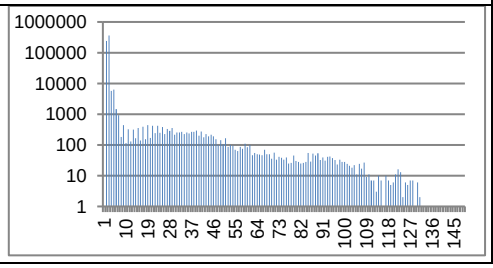
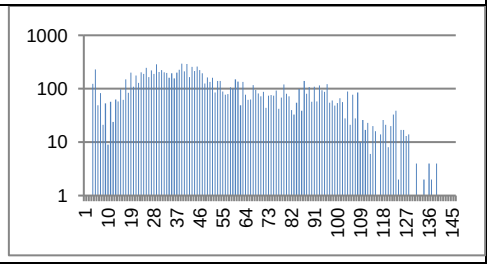
29		
30		
31		
32		
33		
34		
35		

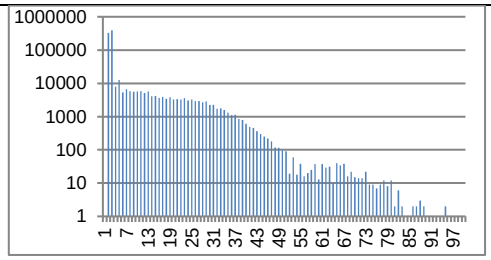
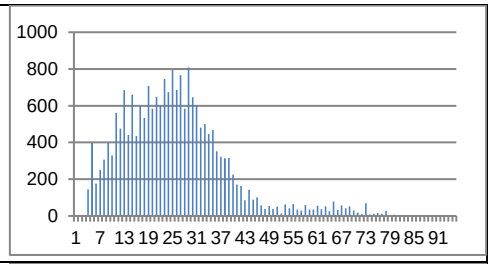
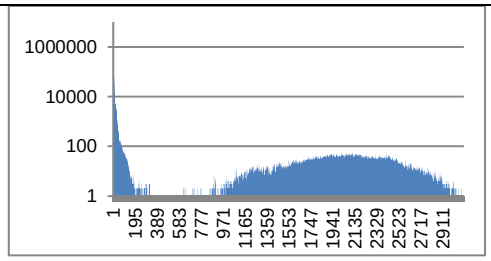
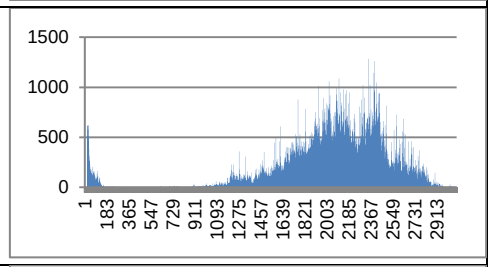
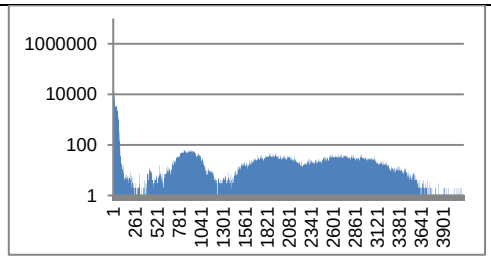
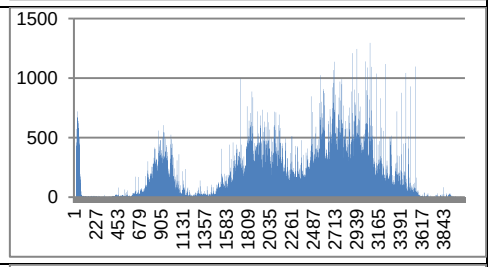
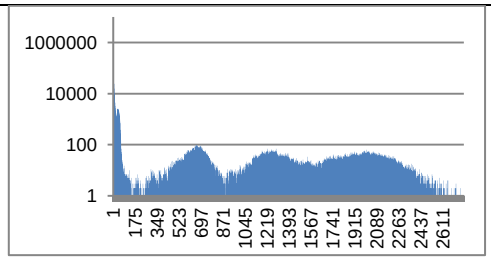
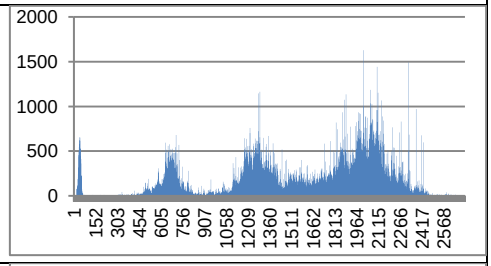
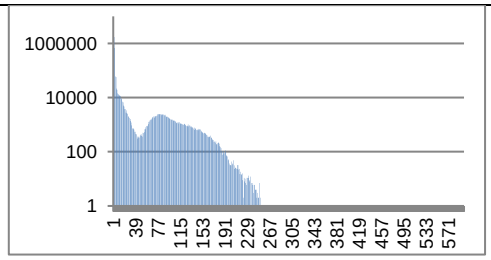
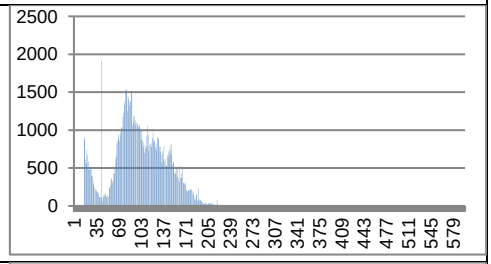
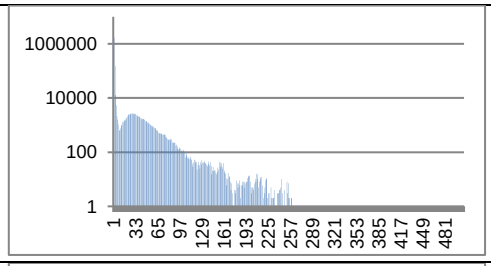
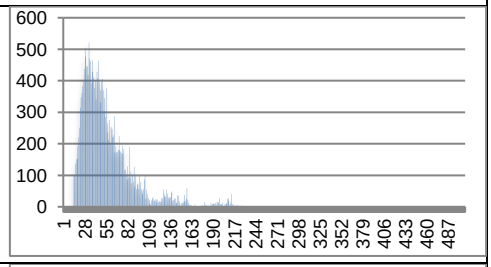
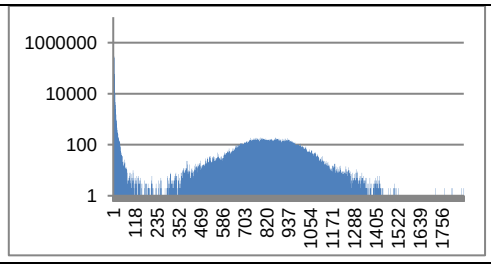
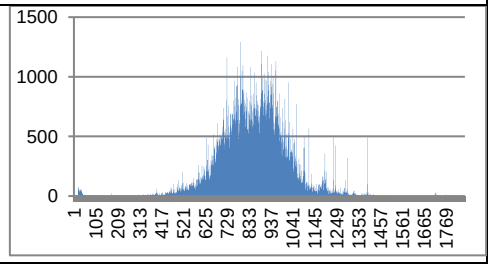
36		
37		
38		
39		
40		
41		
42		

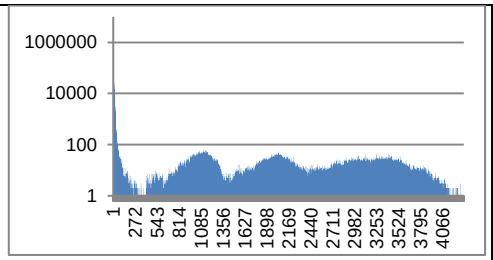
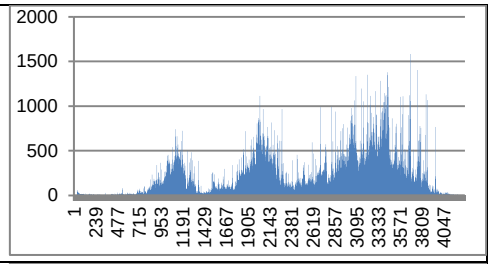
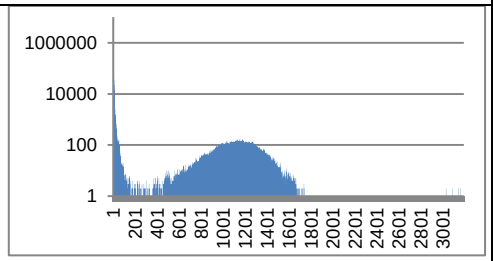
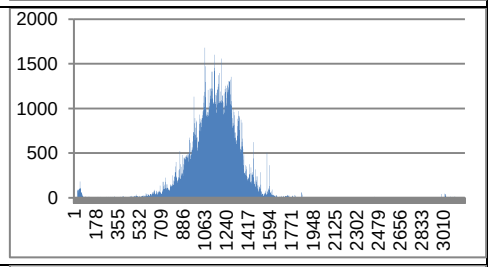
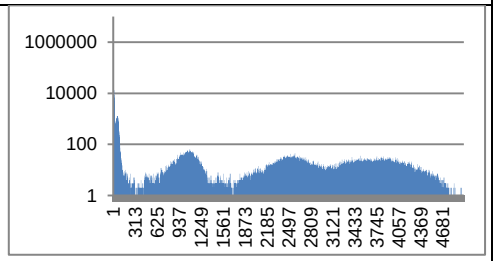
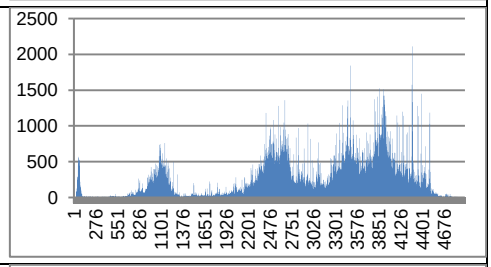
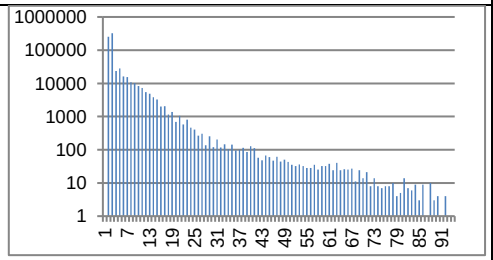
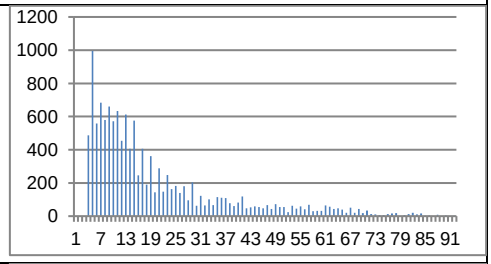
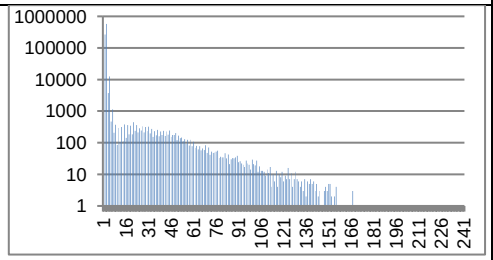
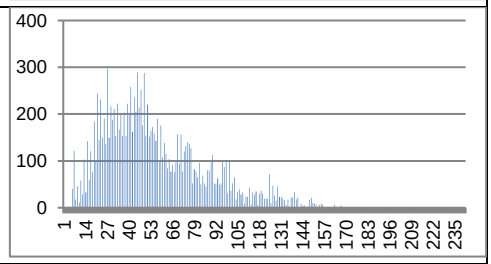
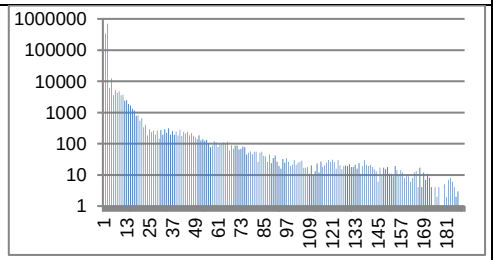
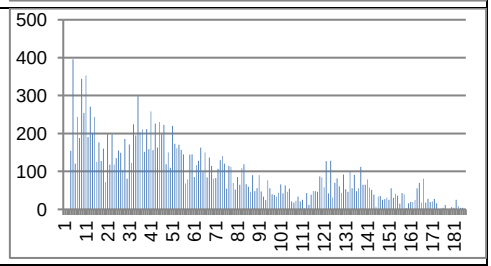
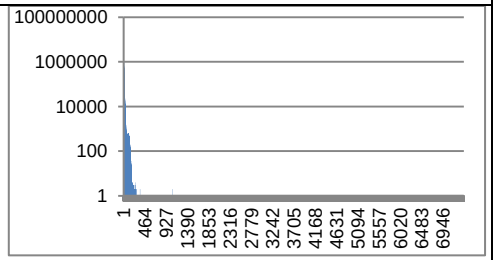
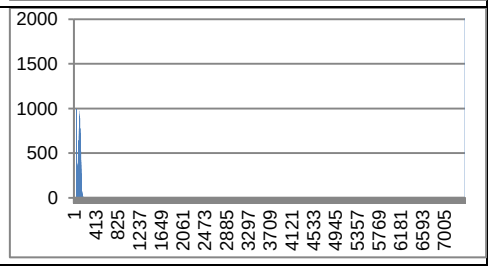
43		
44		
45		
46		
47		
48		
49		

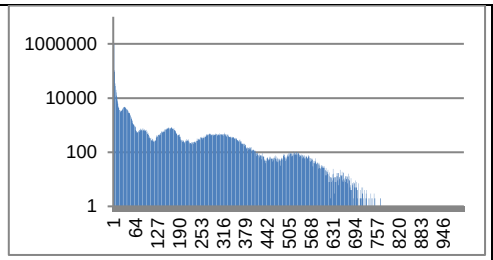
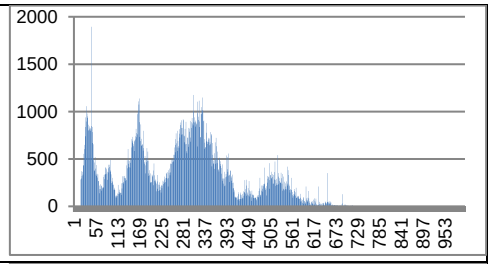
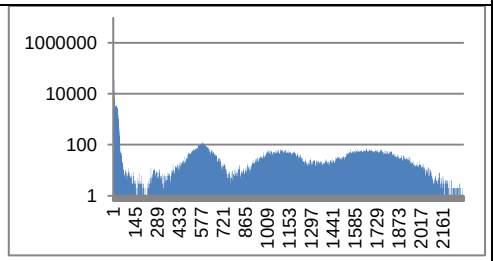
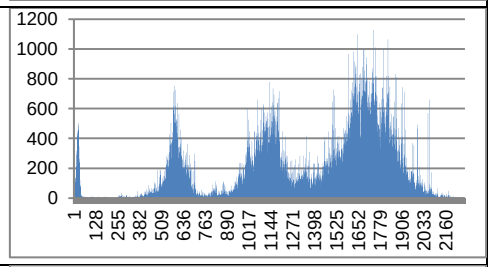
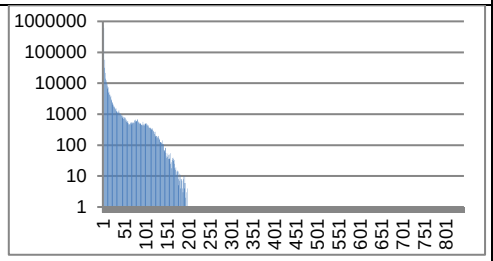
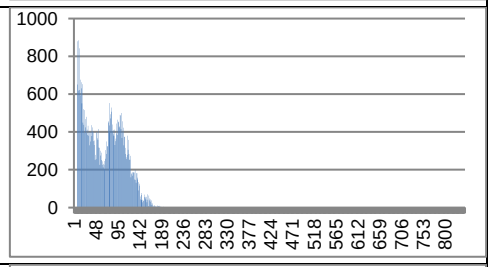
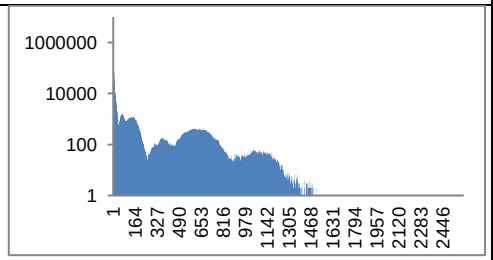
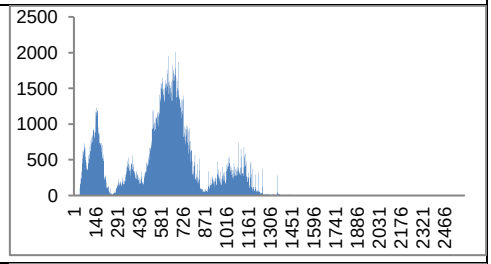
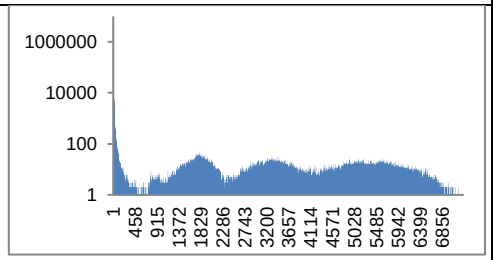
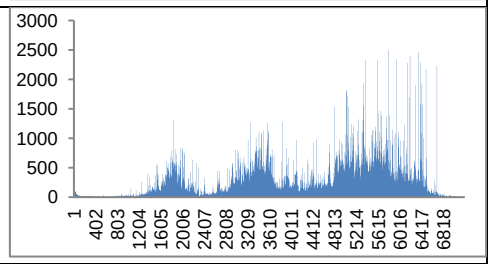
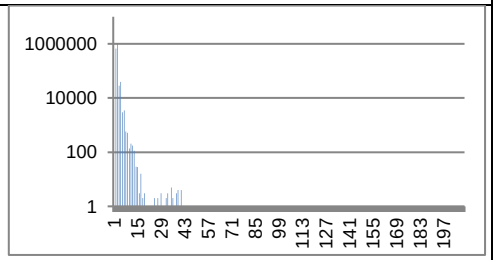
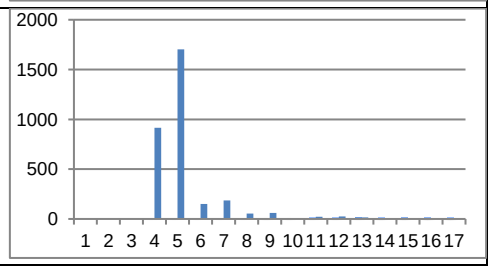
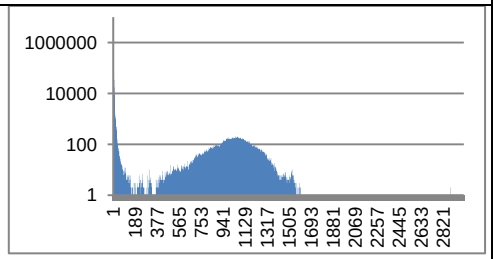
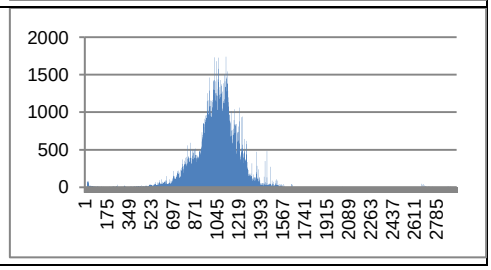
50		
51		
52		
53		
54		
55		
56		

57		
58		
59		
60		
61		
62		
63		

64		
65		
66		Failed
67		
68		
69		
70		

71		
72		
73		
74		
75		
76		
77		

78		
79		
80		
81		
82		
83		
84		

85		
86		
87		
88		
89		
90		
91		

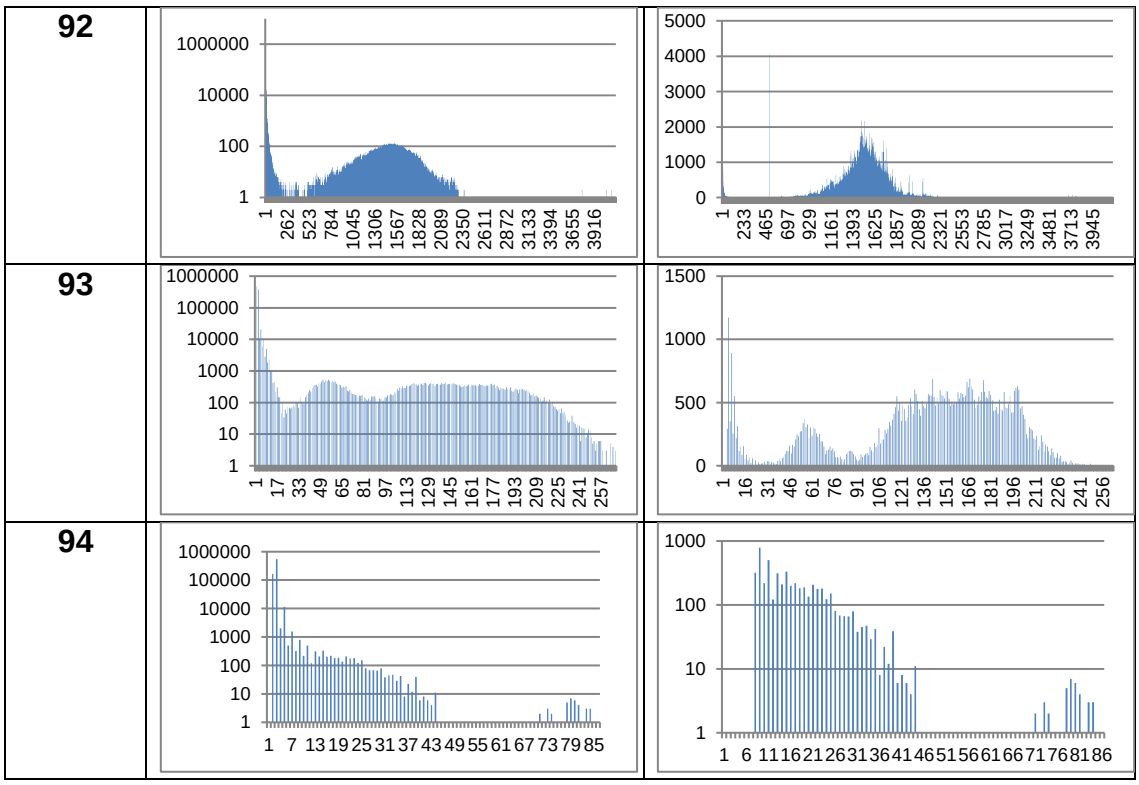


Table 4.1: Pa K-mer peaks pre and post sequence extraction. The table shows the pre and post k-mer graphs. The post sequence extraction graphs illustrate k-mers after removal of low abundance k-mers.

From the 94 samples after preparation and sequencing 107 Pa phages were chosen from the samples which had both good coverage > 10 and were seen to be non-chimeric. These phages were annotated and presented in this section. Two Pa samples 66 and 90 had no assemblies and 21 samples yielded low sequence data for the assembler to give good partial or complete phages. The samples that contained multiple phages and that were initially thought to be representative as separate peaks were assembled separately and if incomplete assemblies were seen then sequence data was merged from the peaks to help improve the assemblies. The evaluation of the mixed phage assemblies exhibiting non-chimeric assemblies were supported using mapping the raw reads back to the assembly contigs and comparison searching using blastn results.

For example, in the sequence data 20 we have 2 k-mer peaks table 4.1 the assembly of the 2 sequence peaks together (k-mer peak1 and peak 2) yields a larger contig around 61.5 kbp when compared to the individual peak 2 assembly 52.6kbp (appendix 2) however, when these reads are mapped back the sequence coverage drops in the last 9 kbp suggesting false elongation of the contig. The sequences in peak two also contain another phage which can be seen in the slight rise in the sequence count in the k-mer graph of sequence sample 20. Figure 4.2 shows an Artemis image of the mapped reads bam reads to node1 (61.5 kbp) and node 2 (37.6 kbp). Therefore the assembly read negating the sequence data from k-mer peak1 and only containing the sequence data in peak 2 was used and presented as the phage assemblies. This example shows firstly why analysing the sequence depth is essential in omitting chimeric assemblies. Sequence sample 20 also shows that the two phages Phi297 and D3112 are genomically different and therefore it was possible to separate them apart from potentially one sequence k-mer peak.

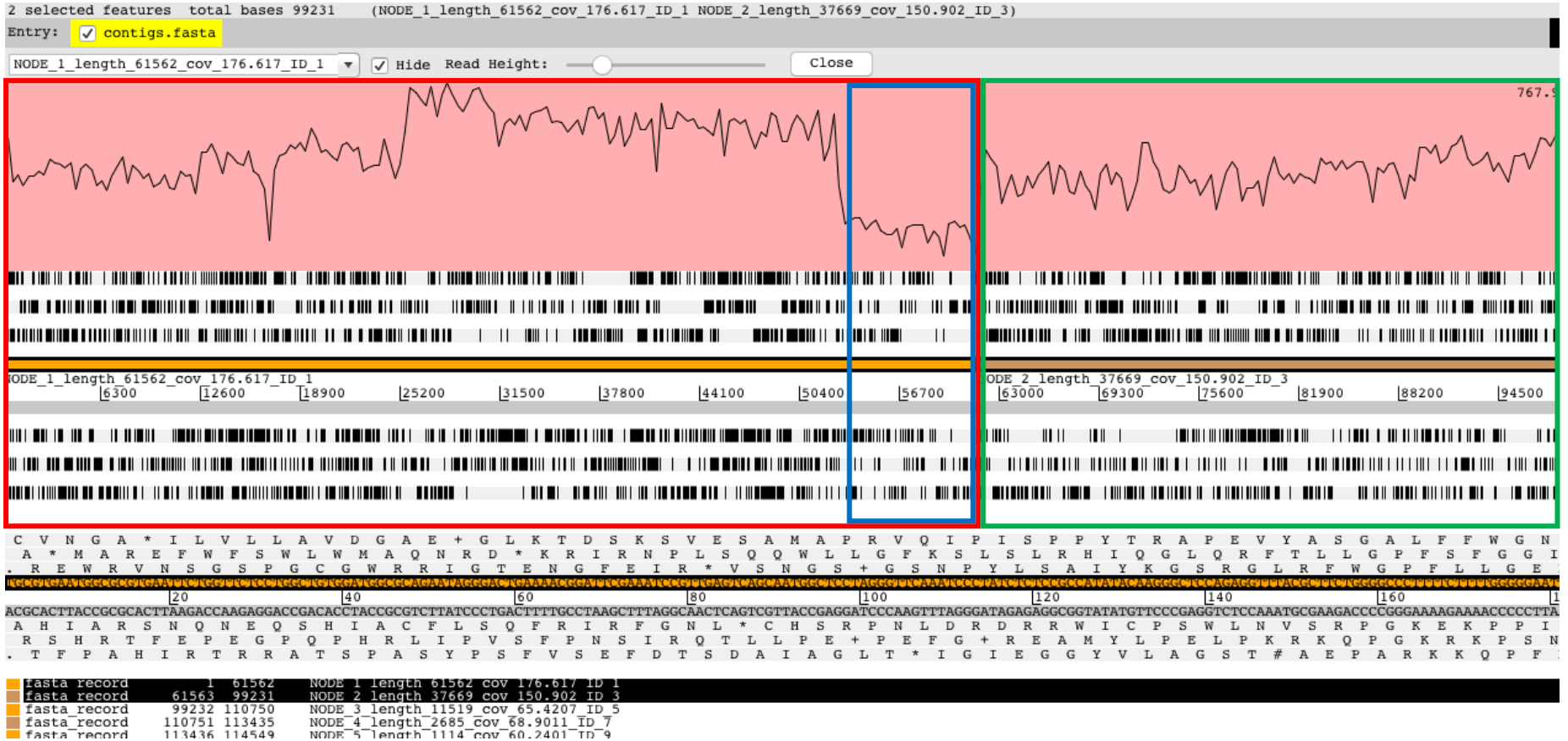


Figure 4.2 Artemis bam reads mapped back to contigs. Sequence sample 20, potentially 2 phages node 1 61.5 kbp in the red rectangle and node 2 37.6 kbp in the green rectangle. However, the sequence coverage drops for the last 9 kbp on node 1 highlighted by the blue rectangle indicating possible forced elongation of the phage.

This sequence data set also highlights the importance of removing sequence data and normalising the reads based on k-mer count and abundance. A table of all the SPAdes assemblies can be found in appendix 2 which were used to determine the best phage assemblies. Due to the large number of phages presented in this study it would be impractical to present those in the main body of the report and it would not be practical to discuss all the assembly evaluation in detail thus these phages with their table of putative functional genes and genome map can be found in appendix 3. To compare the phages a dot matrix graph was generated using gepard (Krumsiek et al, 2007), figures 4.3-5 show this based on the phages grouped to the disease aetiologies adult CF, Paediatric CF, > 10 years BR and < 10 years BR as described in chapter 2.5.1. The greyer the background the more similar the sequence between the phages. A dark black diagonal line represents a homologous region between the phages. The dot matrices show that there are phages which are highly comparable at the genome level, which we could expect to see given that they are all induced from the same bacterial species. However, we still see regions of sequence dissimilarities between the phages. We can also gauge that the sequence orientation changes between the assemblies.

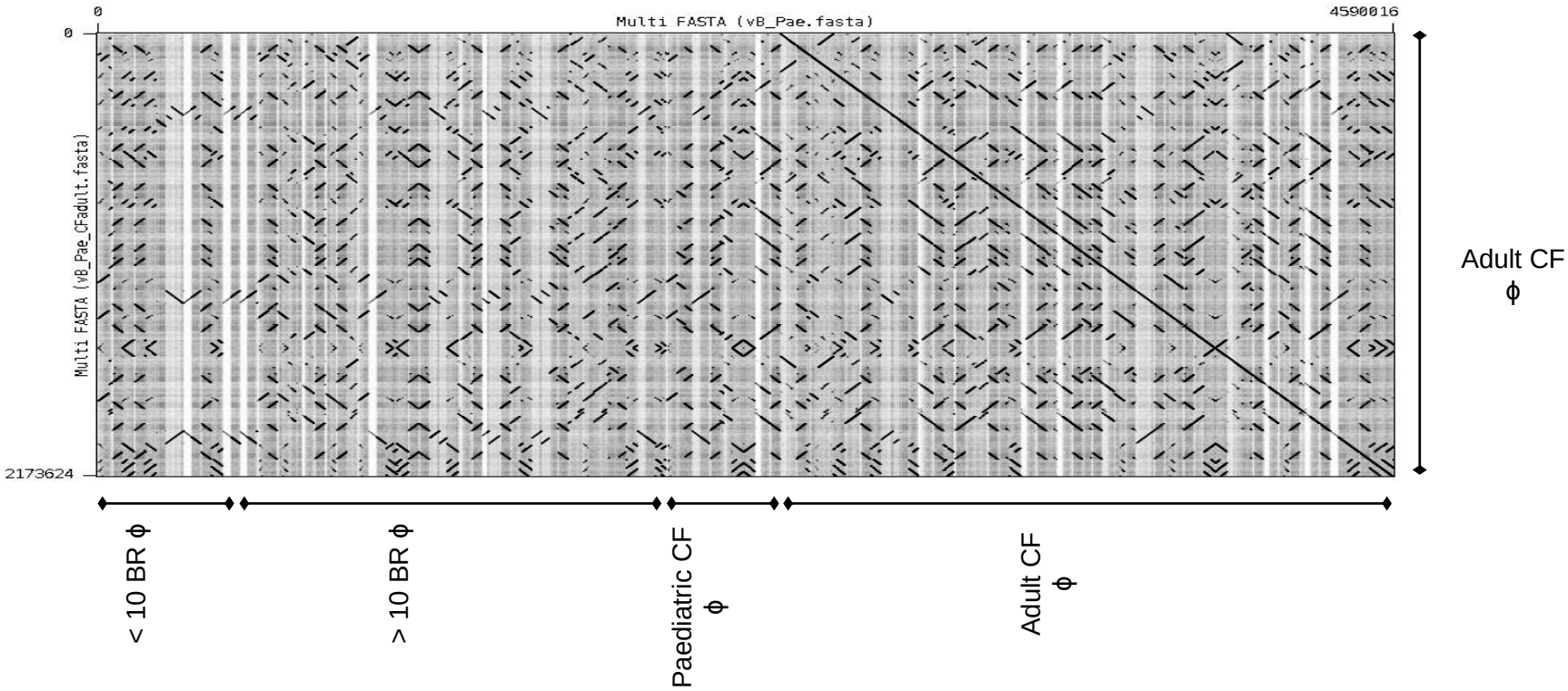


Figure 4.3: Dot matrix of adult CF phage compared to cohort. A dot matrix comparing the nucleotide sequence of the 52 adult CF *P. aeruginosa* phages. The darker the background is the more similar the sequences. The phages were grouped back to the disease aetiologies used previously < 10 years BR, < 10 years BR, Paediatric CF and Adult CF on the x axis and the adult CF on the y axis. This map was generated using Gepard v1.40 (Krumsiek et al., 2007) using a window size of 25 and word length 10.

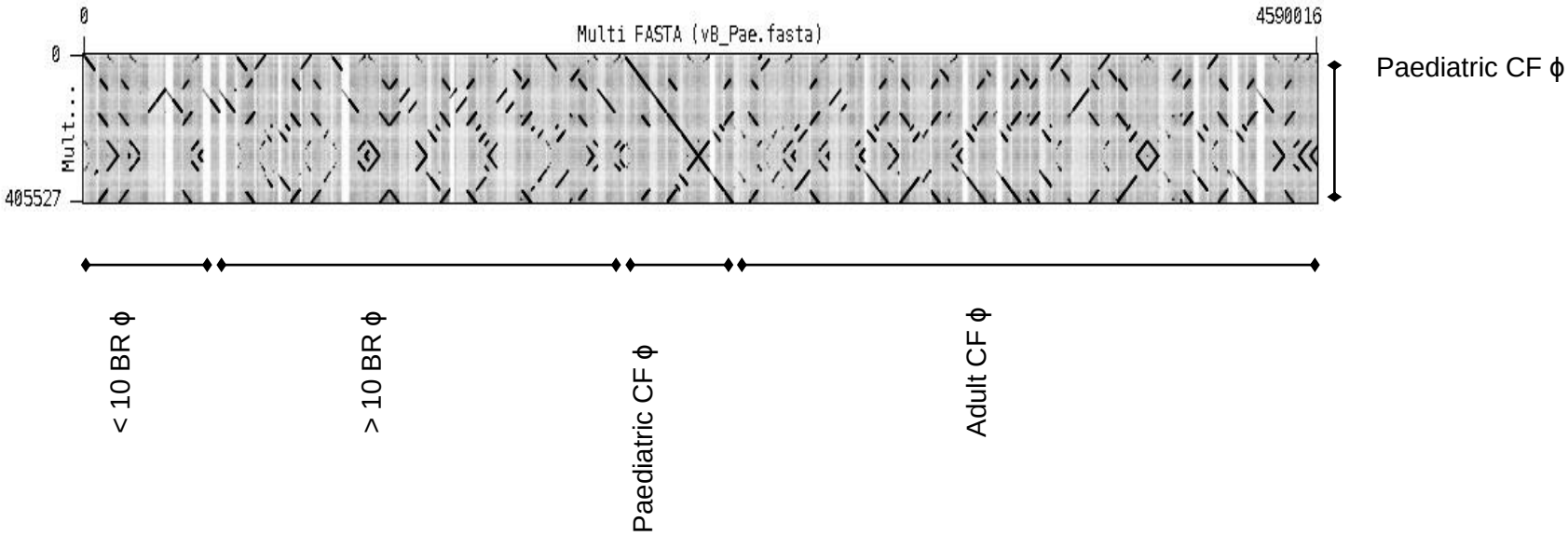


Figure 4.4: Dot matrix of paediatric CF phage compared to cohort. A dot matrix that compares the nucleotide sequence of the 10 Paediatric CF *P. aeruginosa* phages. The darker the background is the more similar the sequences. The phages were grouped back to the disease aetiologies used previously. < 10 years BR, < 10 years BR, Paediatric CF and Adult CF phages on the x axis and the Paediatric CF phages on the y axis. This map was generated using Gepard v1.40 (Krumsiek et al., 2007) using a window size of 25 and word length 10.

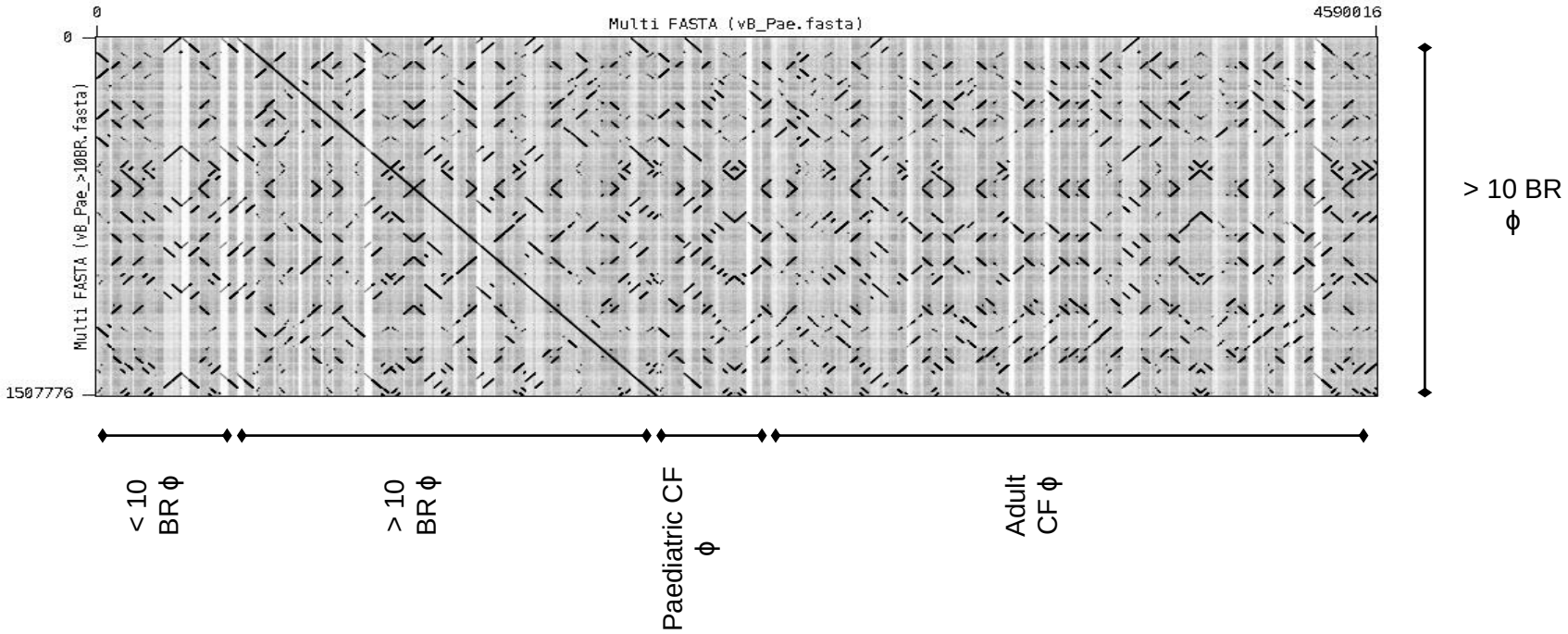


Figure 4.5: Dot matrix of > 10 BR phage compared to cohort A dot matrix that compares the nucleotide sequence of the 35 > 10 years BR *P. aeruginosa* phages. The darker the background is the more similar the sequences. The phages were grouped back to the disease aetiologies used previously. < 10 years BR, < 10 years BR, Paediatric CF and Adult CF phages on the x axis and the > 10 years BR phages on the y axis. This map was generated using Gepard v1.40 (Krumsiek et al., 2007) using a window size of 25 and word length 10.

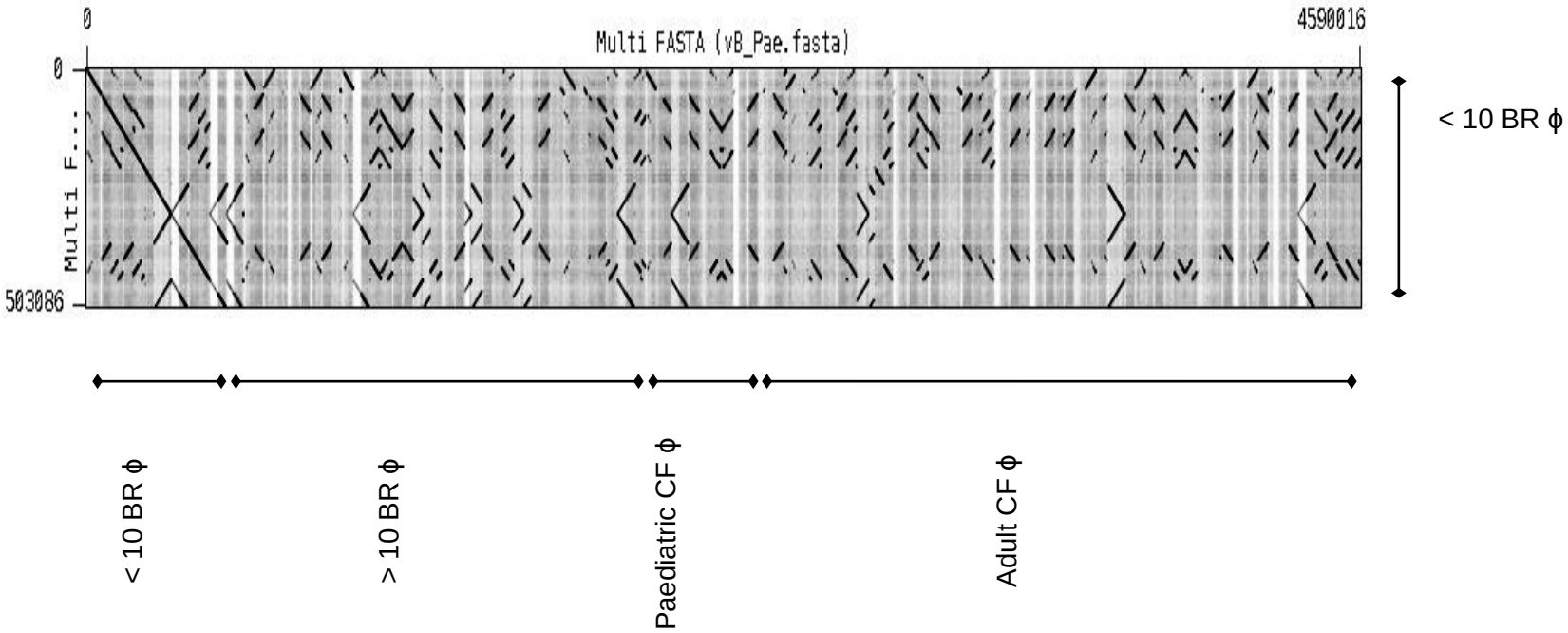


Figure 4.6: Dot matrix of < 10 BR phage compared to cohort A dot matrix that compares the nucleotide sequence of the 11 < 10 years BR *P. aeruginosa* phages. The darker the background is the more similar the sequences. The phages were grouped back to the disease aetiologies used previously. < 10 years BR, < 10 years BR, Paediatrics CF and Adult CF phages on the x axis and the < 10 years BR phages on the y axis. This map was generated using Gepard v1.40 (Krumsiek et al., 2007) using a window size of 25 and word length 10.

Table 4.2 simplifies the blast output for the Pa phages and groups them based on their shared sequence homology. However, the sequence similarity is for the largest segment of sequence similarity and does not take into consideration the whole sequence. For this reason, the representative phages were aligned using Mauve. Figure 4.7-15 show the phages selected to represent the group as per comparison purposes. Analysis of the phages similar to F10, showed four major variations and thus these were used in Figure 4.7 for sequence alignment comparison. Similarly, it was noted when looking at the Pa phage genomes (appendix 3) that there we two major variations of vB_PaeS_PMG1, D3112 and three for phages similar to Phi297 (table 4.2. One representative of H66, JBD24, B3, LKA5 and PhiCTX were used (table 4.2). There was only one type of LKA6 and PhiCTX-like phage (table 4.2). For JBD24 and B3 the assembly with the most repetition and an assembly with a sequence start upstream of an identified terminase gene (if applicable) were used. In the case of phages similar to H66, 11 of the 12 had the assembly length of 61772bp thus a representative of the group was taken. It can be seen from the mauve alignment is that the sequence shares more similarity than the basic blast output data. Looking at the mauve alignments blocks of sequence similarities can be seen represented as different colours.

Figure 4.7 is an alignment comparison between F10 and four phages which showed some sequence similarity to it. F10 (accession: DQ163912.1) is a Pa phage first published in 2006 (Kwan et al, 2006). vB_Pae_CF31c is the least similar and the most similarity can be seen on the right most of the phage genome. The genome of vB_Pae_CF31c begins with a phage terminase. vB_Pae_CF37a and vB_Pae_CF7b share the most sequence homology to F10 however, vary in the genome architecture. vB_Pae_CF37a begins with a Nu1 and Phage terminase gene unlike vB_Pae_CF7b. vB_Pae_BR93a is most similar in the red block but varies along the rest of its genome when compare to F10.

Most similar phage on blast database	Genome size bp	Region size with similarity in kbp	Number of phages similar to the Phage
F10	39199	1.9, 6.2 and 9.7	32
vB_PaeS_PMG1	54024	4.9 and 6.3	6
D3	56425	3.3	1
Phi297	49135	8.6 and 8.8	16
D3112	37611	17.4	20
H66	65270	11.7	12
JBD24	37065	27.1	9
B3	38439	7.1	9
LKA5	64746	13	1
PhiCTX	35580	9.4	1

Table 4.2: Phage diversity and closest genomic member. The table shows the blast similarities output on a viruses' database. The table shows the range in sequence similarities seen and the number of phages showing that similarity. The total numbers of Pa phages tally up to 107.

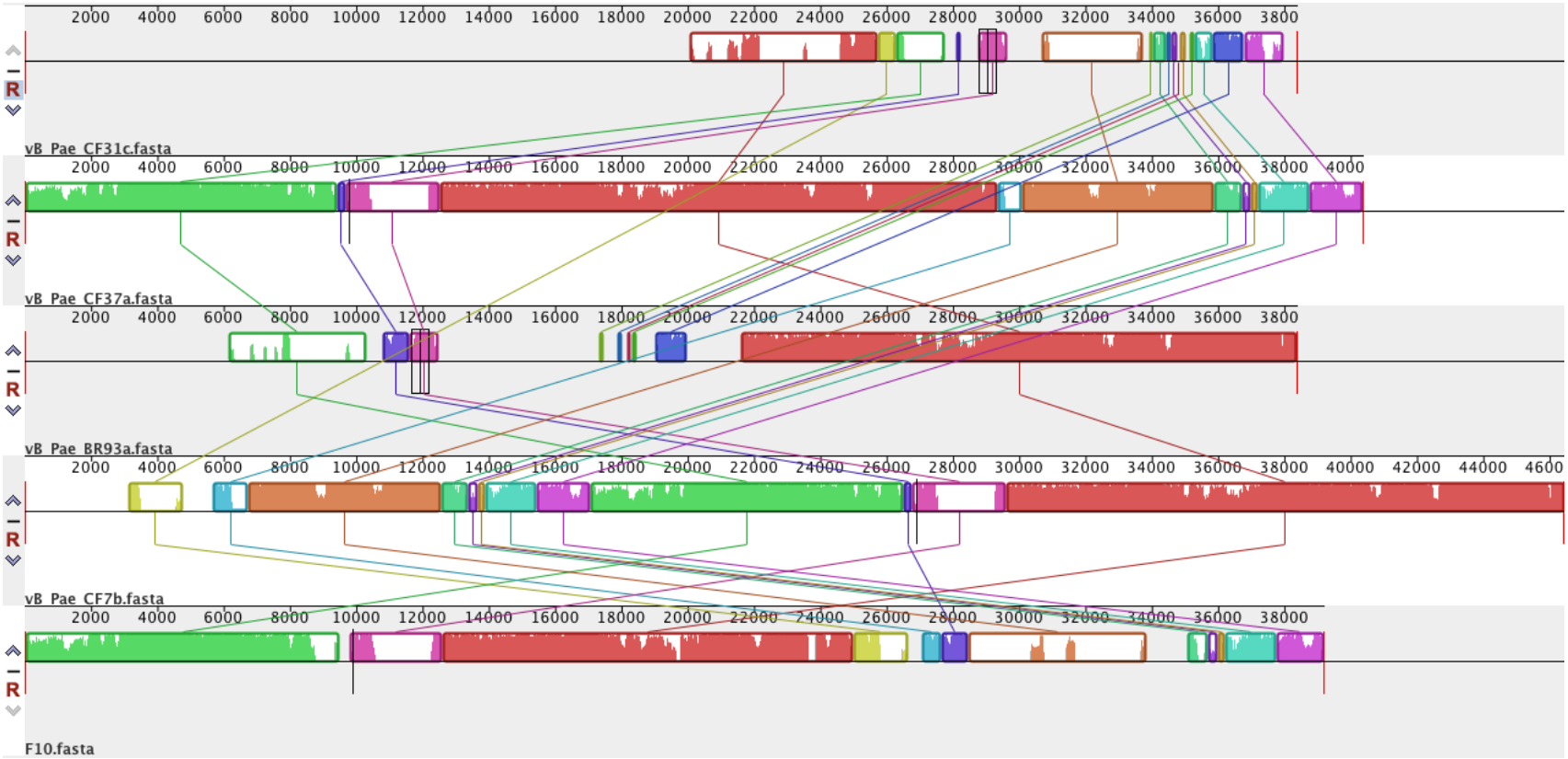


Figure 4.7: Mauve comparison of F10-like phages. A sequence alignment that compares the nucleotide sequence of 4 F10-like (named in the bottom left corner of the image) *P. aeruginosa* phages with F10. The colour backbone makes it easier to see where there are shared sequence matches. This alignment was generated using Mauve v2.3.1 using progressive mauve.

Figure 4.8 compares the alignment of two phages similar to vB_PaeS_PMG1. vB_PaeS_PMG1 sequence was first published in 2010 as a direct sequence submission with the accession HQ711985. vB_Pae_CF16a seems to have a similar genomic architecture between the references whereas the green block at the end of the reference vB_PaeS_PMG1 genome seems to be in the left of the genome architecture of the vB_Pae_BR85c phage. The ORF at around 22 kbp on the reference PMG1 phage is an O-polysaccharide acetyltransferease and O antigen related genes are not present in the two phages shown in this comparison. The integrase gene however at 28 kbp region is reasonably conserved between the phages. All other ORF are hypothetical proteins.

Figure 4.9 looks at three phages that are similar to the Phi297 phage and all are comparable at a similar level however vB_Pae_CF5a seems to confer to the genome architecture of Phi297. Phi297 sequence was first published in 2010 as a direct sequence submission with the accession HQ711984. vB_Pae_BR88b has a genomic sequence beginning with a terminase and is relatively smaller in size to the other phages in this alignment. The red block codes for phage morphogenesis, phage coat proteins and tail genes. The gap in the reference Phi297 genome at 15 – 21 Kbp entails phage repressor, replicative DNA helicase, ParA protein and DNA replication protein O.

Figure 4.10 genomically compares D3112 to two phages that were similar based on sequence homology to D3112 but varied in sequence length. D3112 was first sequenced in 2004 with the accession AY394005 and was identified as a transposable Pa bacteriophage (Wang et al, 2004). vB_Pae_CF25a is 4 Kbp larger on the left and around 2 Kbp larger on the right and shares no sequence homology to either the other phage or the reference D3112. The ORF in the leftward genome of vB_Pae_CF25a are IgA, RecT and YqaJ-like viral recombinase and on the rightward side codes for a hypothetical quiP gene. There are 4 clear regions in the reference D3112 genome where there is no sequence homology to any other induced phage. These ORFs noted are all hypothetical and offer no functional information.

Figure 4.11 compares a *Podoviridae* H66 phage to a phage sharing 11.7 kbp sequence homology based on blastn output. Pa phage H66 was directly submitted to the sequence database with the accession KC262634 in 2012. The phage vB_Pae_BR59a is used as an example however; all have the same sequence assembly exactly to 61772bp length. The red block in the vB_Pae_BR59a regions has no sequence homology to the reference H66 genome and code for DNA methylase genes, the gap between the red and green block at around 15 kbp is situated by an integrase gene in the vB_Pae_BR59a. The large green block of sequence homology has ORF described as hypothetical proteins.

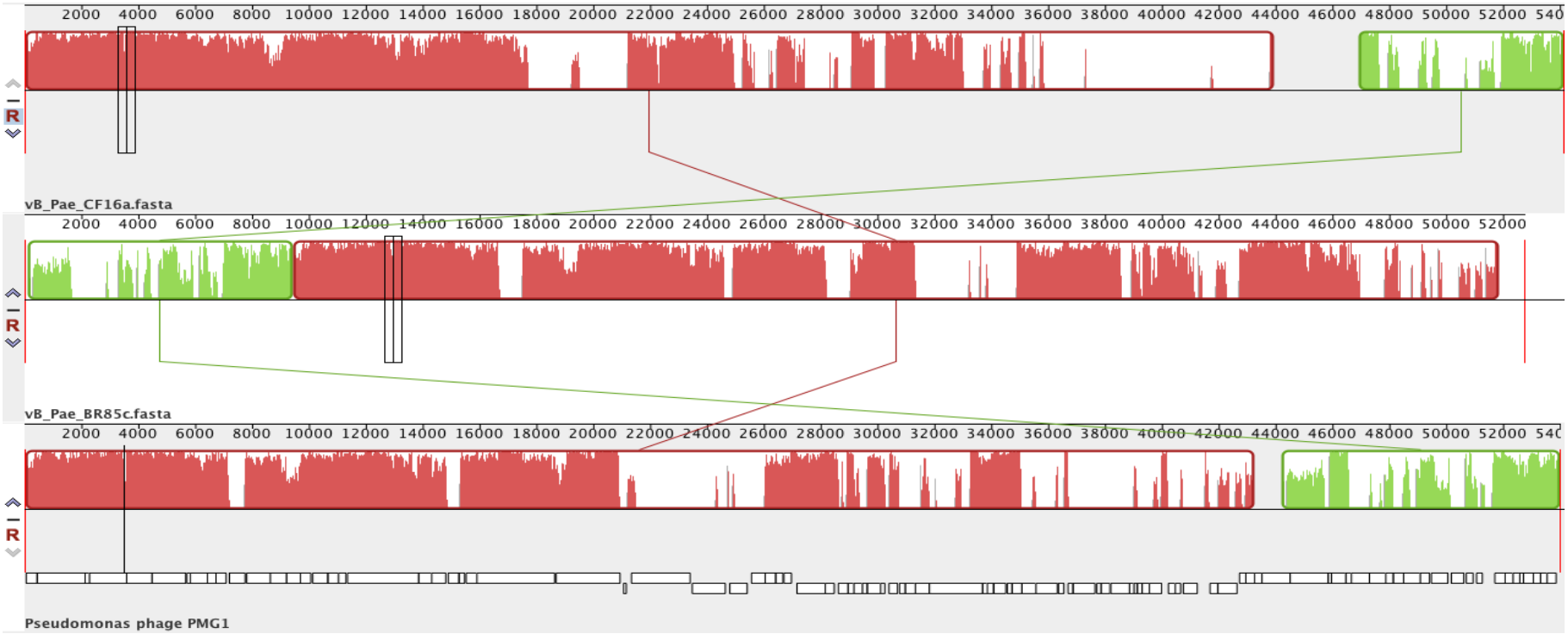


Figure 4.8: Mauve comparison PMG1-like phages. A sequence alignment that compares the nucleotide sequence of 2 vB_PaeS_PMG1-like (named in the bottom left corner of the image) *P. aeruginosa* phages with vB_PaeS_PMG1. The colour backbone makes it easier to see where there are shared sequence matches. This alignment was generated using Mauve v2.3.1 using progressive mauve.

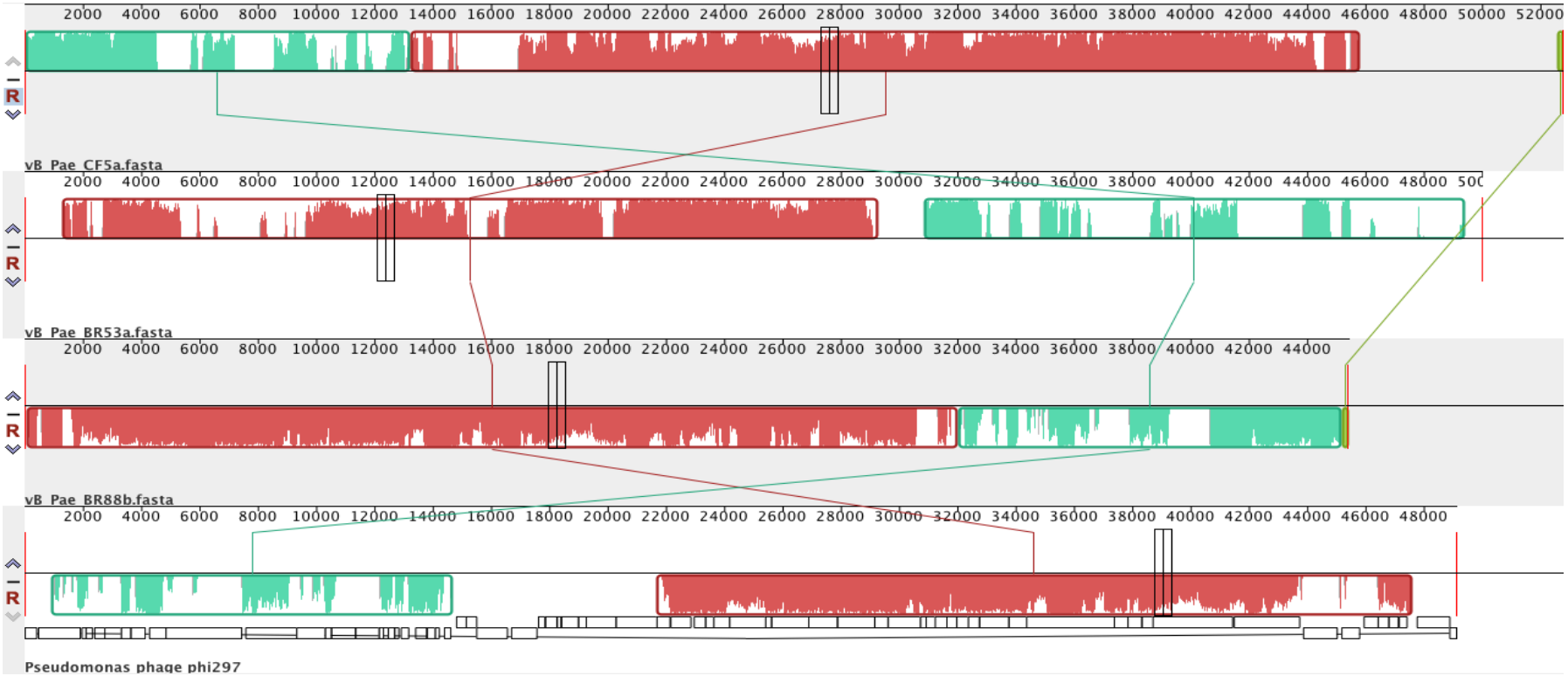


Figure 4.9: Mauve comparison of Phi297-like phages. A sequence alignment that compares the nucleotide sequence of 2 Phi297-like (named in the bottom left corner of the image) *P. aeruginosa* phages with Phi297. The colour backbone makes it easier to see where there are shared sequence matches. This alignment was generated using Mauve v2.3.1 using progressive mauve.

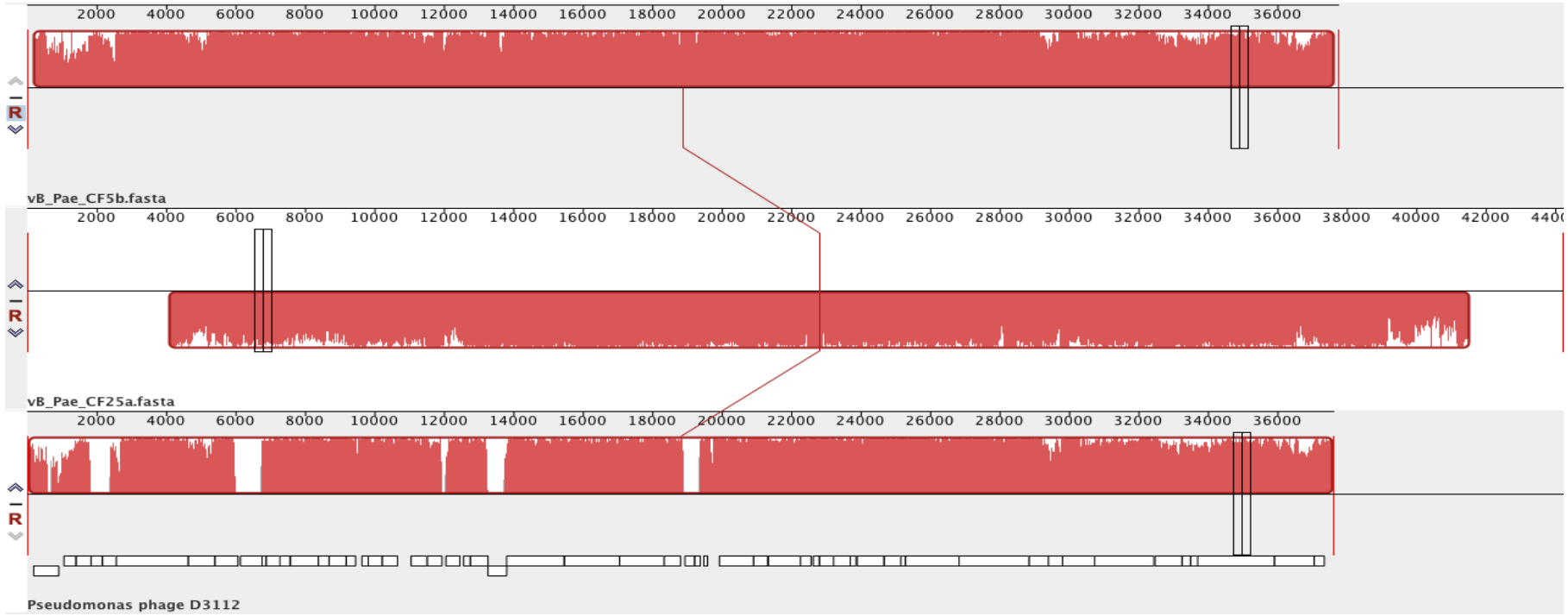


Figure 4.10: Mauve comparison of D3112-like phages. A sequence alignment that compares the nucleotide sequence of 2 D3112-like (named in the bottom left corner of the image) *P. aeruginosa* phages with D3112. The colour backbone makes it easier to see where there are shared sequence matches. This alignment was generated using Mauve v2.3.1 using progressive mauve.

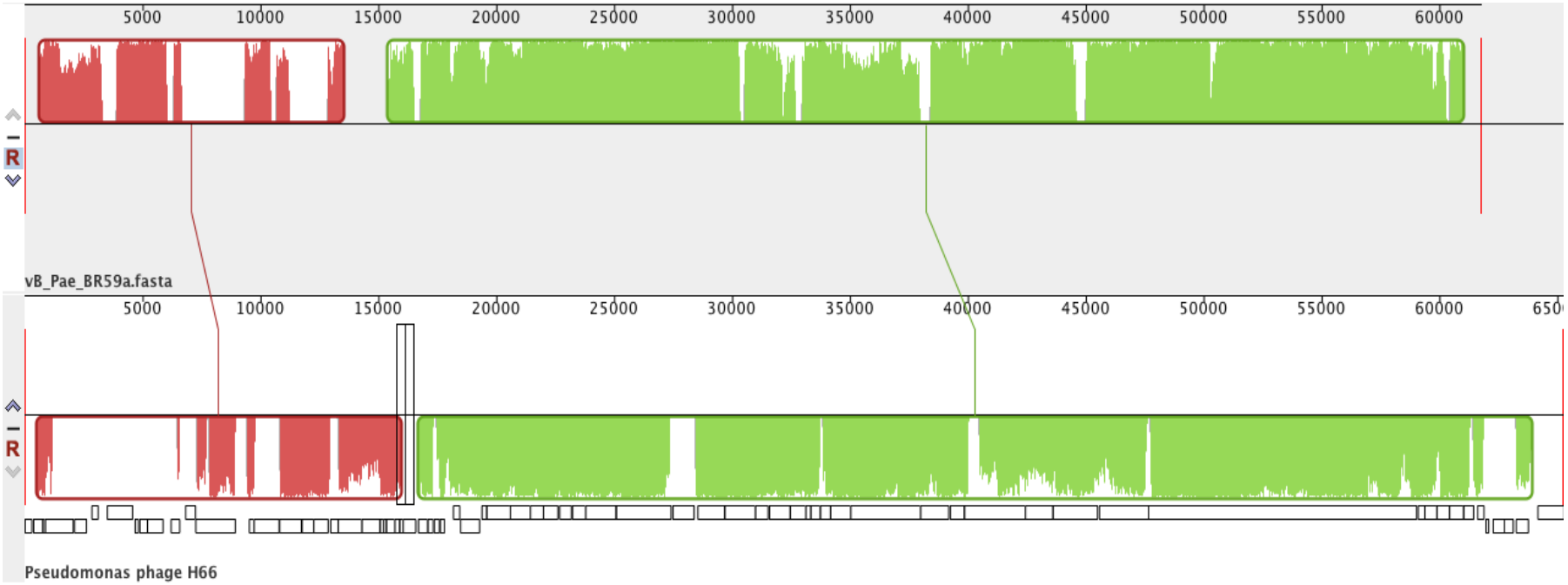


Figure 4.11: Mauve comparison of H66-like phages. A sequence alignment that compares the nucleotide sequence of an H66-like (named in the bottom left corner of the image) *P. aeruginosa* phage with H66. The colour backbone makes it easier to see where there are shared sequence matches. This alignment was generated using Mauve v2.3.1 using progressive mauve.

Figures 4.12 aligns JBD24 with vB_Pae_CF7a which shares sequence homology for 27.1kbp however, this similarity is shared between 10-30,000bp of the phage genome. Indicating most variations in the outer regions of the genome but predominant variation is seen on the left of the genome. The ORF that share no homology are hypothetical proteins. JBD24 was first sequenced in 2013 with the accession NC_020203 and identified as having 5 genes that inactivate the bacterial CRISPR/CAS system (Bondy-Denomy et al, 2013).

Figure 4.13 looks at B3 compared to vB_Pae_CF27a. B3 is a Mu-like transposable phage first sequenced in 2004 with the accession AF232233 (Braid et al, 2004). The shared homology is 7.1 kbp which seems to be in the right hand side of the genome. There is shared similarity which is dispersed across the whole genome. In the reference genome at around the 15 kbp region a large terminase subunit gene is seen which is absent in vB_Pae_CF27a. The same occurs at the 26Kbp region which is a DNA methylase gene. Other genes that don't share homology are hypothetical.

Figure 4.14 shows the comparison between LKA5 and vB_Pae_CF19a. LKA5 was directly submitted to the sequence database with the accession KC900378 in 2013. The sequence homology is 13 kbp however, there is scattered sequence homology across the genome and the sequence assembly is seen to be inverted from the centre. The phages are similar in their DNA methylase, DNA specific binding, NinG, portal, capsid and the integrase gene. However, dissimilarities are seen but all are hypothetical protein genes.

Figure 4.15 compares PhiCTX with vB_Pae_BR76a sequence. PhiCTX was first submitted in 1989 with the accession AB008550 and was shown to encode for a cytotoxin gene (ctx) (Hayashi et al, 1989). The sequence is reversed in comparison to PhiCTX but importantly 2 regions of dissimilarities separate 3 blocks. The largest region of sequence homology is 9.4 kbp to PhiCTX. The first gene of potential importance is the

CTX gene which is conserved between the two genomes. The regions of dissimilarities are hypothetical proteins.

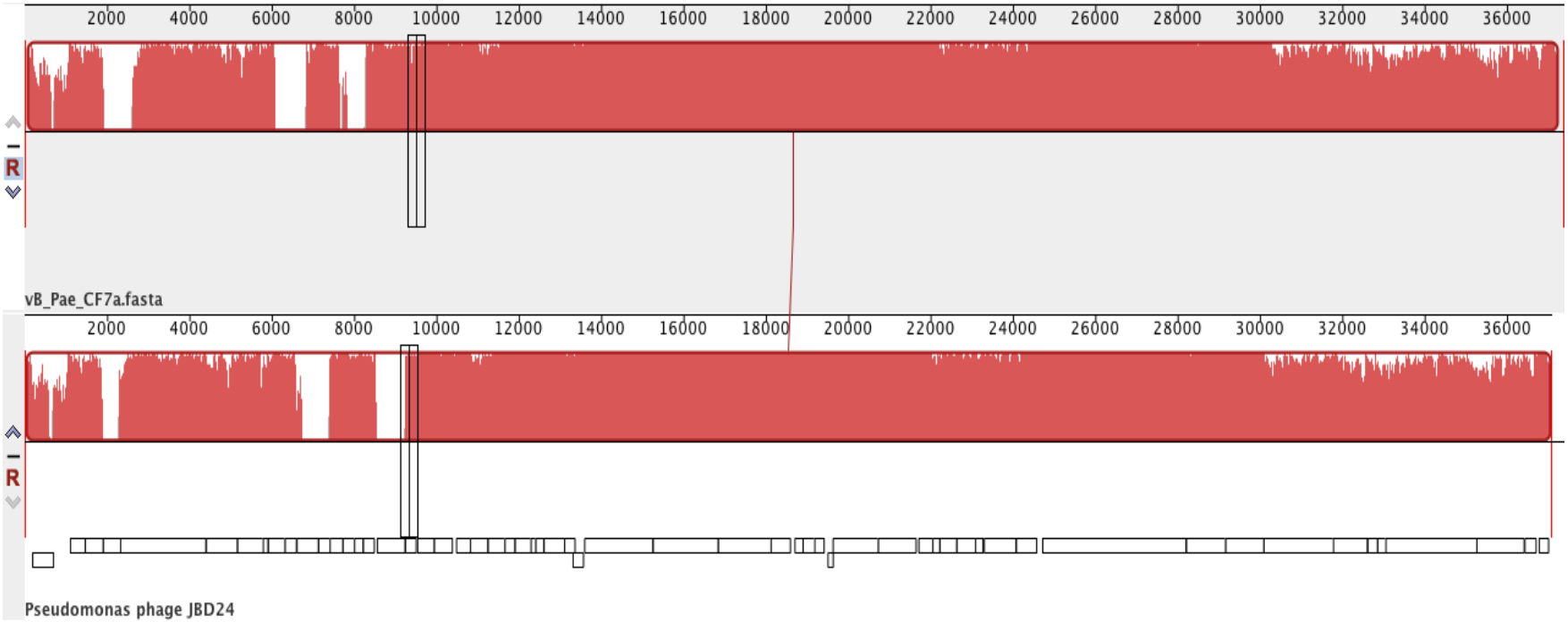


Figure 4.12: Mauve comparison of JBD24-like phages. A sequence alignment that compares the nucleotide sequence of an JBD24-like (named in the bottom left corner of the image) *P. aeruginosa* phage with JBD24. The colour backbone makes it easier to see where there are shared sequence matches. This alignment was generated using Mauve v2.3.1 using progressive mauve.

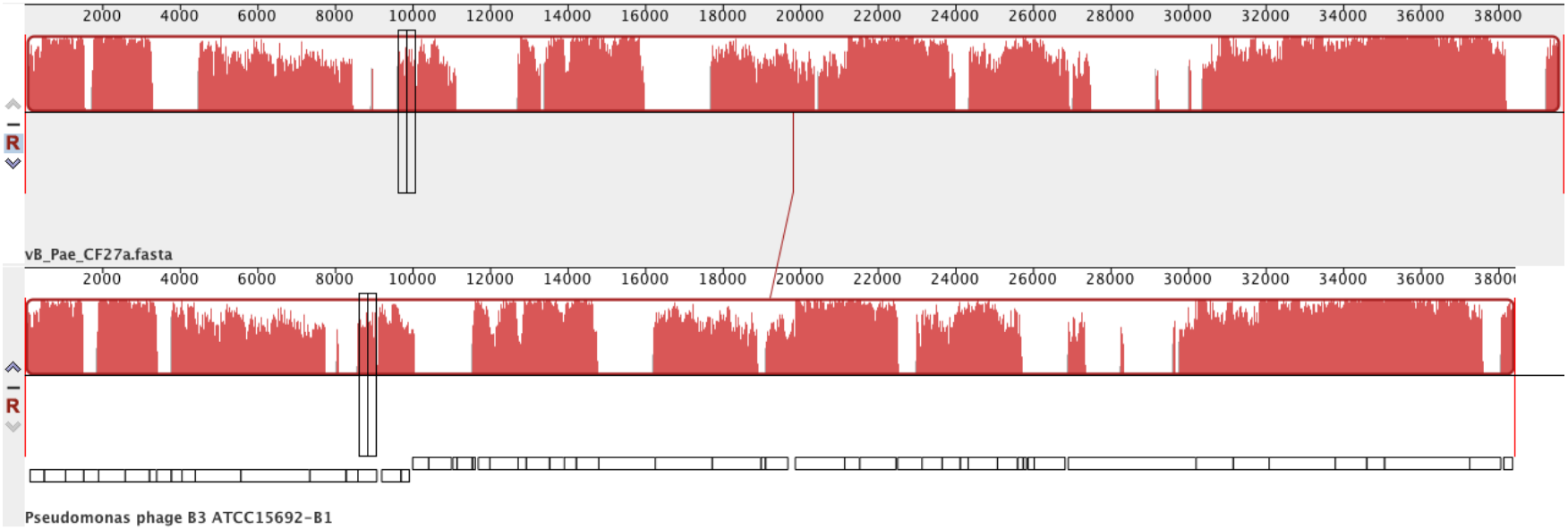


Figure 4.13: Mauve comparison of B3-like phages. A sequence alignment that compares the nucleotide sequence of an B3-like (named in the bottom left corner of the image) *P. aeruginosa* phage with B3. The colour backbone makes it easier to see where there are shared sequence matches. This alignment was generated using Mauve v2.3.1 using progressive mauve.

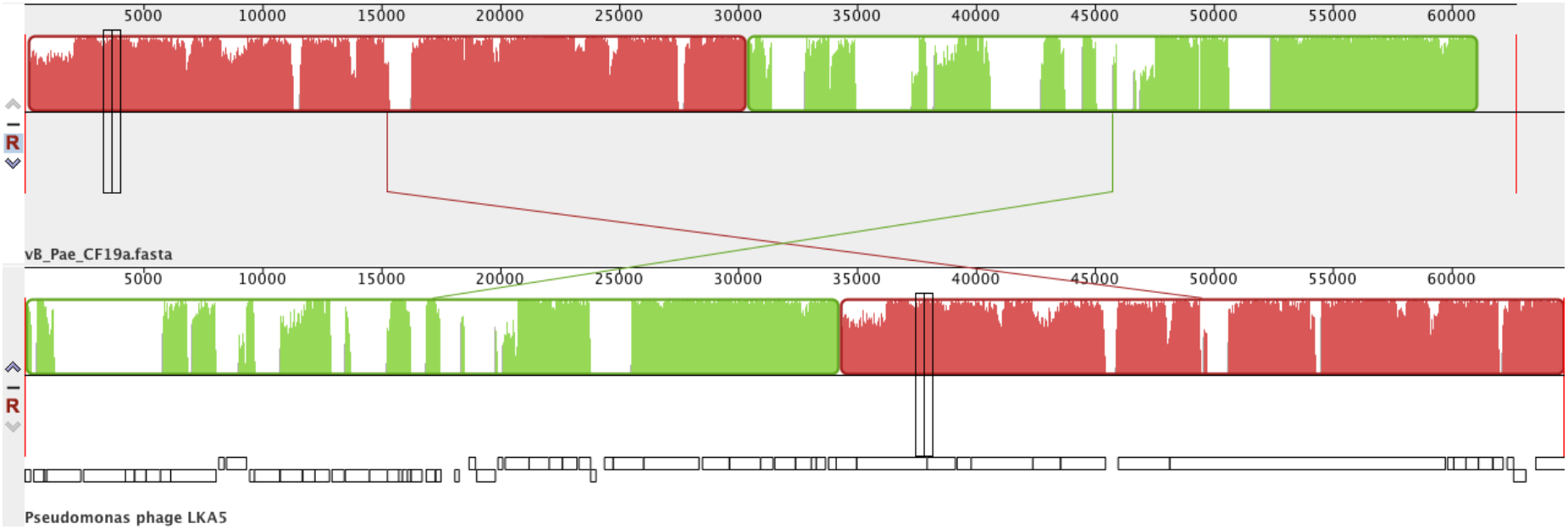


Figure 4.14: Mauve comparison of LKA5-like phages. A sequence alignment that compares the nucleotide sequence of an LKA5-like (named in the bottom left corner of the image) *P. aeruginosa* phage with LKA5. The colour backbone makes it easier to see where there are shared sequence matches. This alignment was generated using Mauve v2.3.1 using progressive mauve.

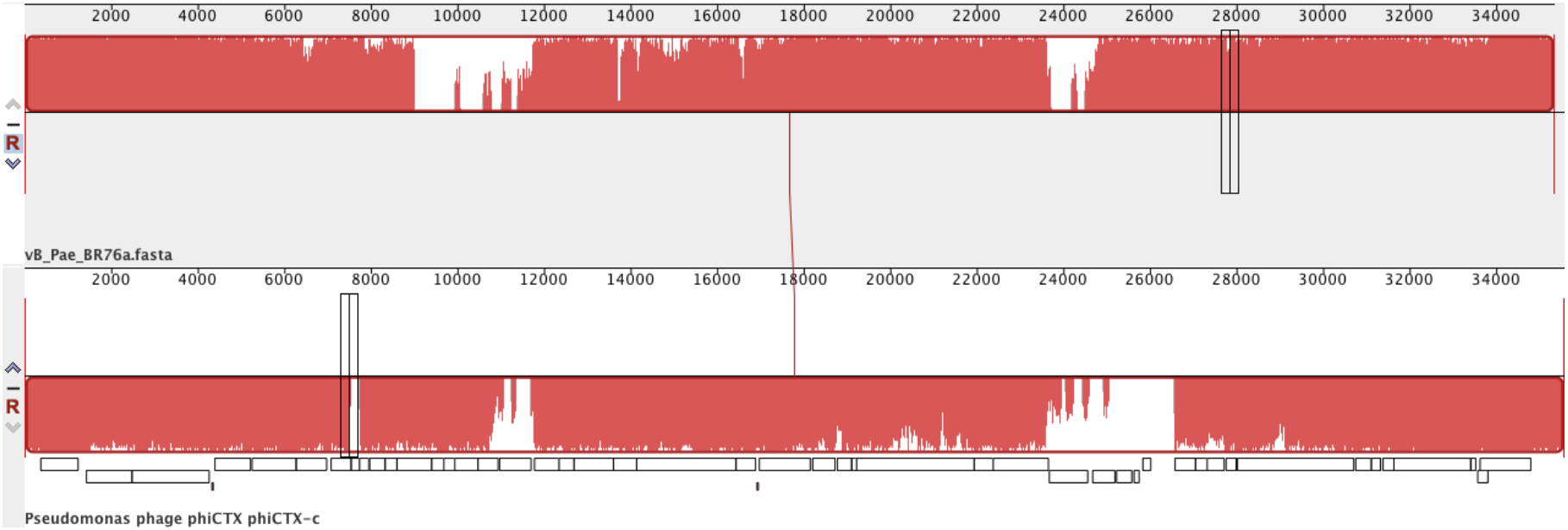


Figure 4.15: Mauve comparison of phiCTX-like phages. A sequence alignment that compares the nucleotide sequence of a PhiCTX-like (named in the bottom left corner of the image) *P. aeruginosa* phage with PhiCTX. The colour backbone makes it easier to see where there are shared sequence matches. This alignment was generated using Mauve v2.3.1 using progressive mauve.

4.3 Discussion

For the purpose of genome sequencing samples were renumbered to 1-94 for simplicity. Table 2.6.2 shows the revised genome sample number corresponding to the Pa sample number in Table 2.2.1 and 2.2.2.

Khmer toolkit was used to remove erroneous k-mers and those associated with contaminating bacterial chromosome (Crusoe et al., 2015). Figure 4.1 shows three examples of how the data was cleaned up of bacterial contamination and k-mers of k length 32 or less removed prior to extracting the sequences in order to assemble the phage using SPAdes *de novo* assembler (v3.5.0). All the k-mer graphs both pre and post sequence extraction can be found in the table 4.1. The “calculate median distribution” script was used and altered to extract the sequences associated with the peaks. These sequence files were assembled separately at first however, it was found that the assemblies were partial and when separate peak sequence files the files were concatenated chimeric assemblies were seen. This was the case when the samples lysate exhibited similar phages and thus it was decided to use different peaks to compare them to see which assemblies gave the highest and more importantly non-chimeric contigs. Figure 4.3 shows an evaluation of the sequence data in sample 20 and from which it was evident that in some instances keeping the peaks separate would help the assembly of the phage genomes. The table of all the assemblies of the Pa phage samples using the different peaks can be found in appendix 2. This approach to assemble phages direct from lysate assembly and removing bacterial chromosomal DNA is a novel method and not described in the literature. We here present therefore mixed phage infection profiles (chapter 3) and the sequence extraction and characterisation of the viral community. From the k-mer graphs we could predict the number of potential phages present in each Pa lysate.

Phages that assembled as multiple contigs, inferring that there may be insufficient sequence depth to bridge the gaps between the contigs, were extended

usingPaired-Read Iterative Contig Extension (PriceTI) (Ruby et al., 2013). PriceTI assembler uses the contigs given to it and maps them back to the raw sequence file with the sole intention to iteratively extend the contigs given to it. The script used for this method can be found in methods 2.8.1. For phage genome where PriceTI recovered or managed to bridge the sequence gaps, were specified in appendix 3.

The 107 Pa temperate phages presented in this study all shared similarities to 10 specific Pa phages that can be found on the NCBI viruses' database. These included *Pseudomonas* phage F10, D3, Phi297, D3112, H66, JBD24, B3, LKA5, PhiCTX and vB_PaeS_PMG1. All these phages are from the *Siphoviridae* class apart from H66 which is a *Podoviridae*. Due to the constraints in time all the phage assemblies were not oriented or broken upstream of the terminase gene manually and were not checked if they circularised at the *cos* sites. The phage assemblies are presented as preliminary drafts and as linear genomes.

The most common similarities were seen to the *Pseudomonas* Phage F10 (Kwan et al, 2006), with 32 temperate Pa phages sharing between 1.9 to 10 kb sequence similarities to this phage. Within the F10 like phages it could be seen that the phages were different. This was perceived to be due to a region in the sequence showing repeat regions of a putative Rha protein. An example where this was seen is *Pseudomonas* phage vB_Pae_CF6b; this gene was predicted twice in the genome. In these phages the *pRha* gene was predicted in the forward and reverse direction. In the reverse strand the *pRha* gene was found close to the late genes and in the forward strand close to the integrase gene. This could imply that one *pRha* is associated with the late genes and the other with phage integration. The pRha protein is a regulatory protein such that it prevents the infection of bacterial strains lacking integration host factor (IHF). IHF is a regulator of the *Rha* gene. vB_Pae_CF7b is an example of a phage that had three *pRha* genes identified. For the first two the same architecture was seen as described above for vB_Pae_CF6b, the additional *pRha* gene was seen down stream of an Arc like DNA binding domain. However, F10-like phages with no repeats of *pRha* were also seen and

had a 38.3 kb size genome and shared only 1.9 kb sequence matches to F10. Hence on a general rule all of the 32 F10-like phages identified in this study would fall in one of the three types of F10-like phages described above. Having a pRha protein could therefore be advantages in preventing infection if the bacterial host did not have IHF as for temperate bacteriophages it is also required to excise. One could assume having multiple copies of this protein could enhance its effect. On the other hand, the protein could be highly conserved and thus an area where the assembler struggles to differentiate between multiple phages. It must also be noted that similarities of up to a maximum of 10 kbp, even though this similarity is important it shows that the remainder of the genome does not compare to any other phage genome on a current database.

Pseudomonas D3112-like (n = 20) (Wang et al, 2004) and Phi297-like (n = 16) (Kropinski et al, 2010a) phages were seen to be present together in 15 Pa sample. In one sample Phi 297 was seen to be present independent of D3112. In this instance the coverage was high and therefore one could assume this reflect the titre of this phage. D3112-like phages in the presence of Phi297-like phages were predominantly seen to have higher sequence coverage. In 3 instances where only D3112-like phages were present in the Pa samples, incomplete fragment similar to the Phi297 phages were seen. In only two samples D3112-like phages were seen independent of Phi297-like phages. However, in those samples the coverage and thus the sequence depth were low. From this it could be assumed that D3112 and Phi297-like phages complement each other as inducible prophage regions. It could be assumed that D3112-like phages are the dominant phage and Phi297-like phages are secondary.

Twelve Pa samples had a phage 61.7 kb length that shared an 11.6 kb sequence similarity to *Pseudomonas* phage H66 (Maya et al, 2012) / or F116 (10.6 kb) (Byrne & Kropinski, 2005). These H66-like phages were always seen to be at 61772bp length and shared the same genome homology. The only variation seen was with the orientation of these phages. The H66 phage is a *Podoviradae* and different genomically to the rest of

the phages. The light bands seen in the dot matrices are predominantly these phages and thus show their genomic diversity against the predominant F10-like phages.

Six of the assembled phages contain a 6.3 kb homologous region to vB_PaeS_PMG1 (Kropinski et al, 2010b). Nine phages had the highest sequence similarity of 27 kb to a *Pseudomonas* phage JBD24 in the viruses' database (Bondy-Denomy et al, 2013). Nine phages shared a sequence similarity to B3 with a low sequence match of 2.5 kb (Braid et al, 2004). One phage had a 13 kb match to *Pseudomonas* phage LKA5 (Lammens & Lavigne, 2013). One phage shared a 9.4 kb match to a *Pseudomonas* phage PhiCTX phage (Nakayama et al, 1999). One phage shared sequence similarity to *Pseudomonas* phage D3 (Kropinski, 2000) although as this was a partial sequence it is most possibly like a Phi297 phage.

Figures 4.3 - 6 showing the dot plot of the bacteriophages separately grouped based on their Pa disease associations (Methods 2.5.1) show a grey scale gradient based on sequence similarity. As described above all of the Pa phages can be inferred to be similar based on sequence similarity to ten Pa phages found in the viruses database. However, clear section of white bands can be seen which indicates that these phages or regions of phage are dissimilar from the overall consensus of phages shown. The region that have the whitest are associated with the H66-like phages this may reflect the fact that H66 is a *Podoviradae* and the other phages share more sequence homology with *Siphoviridae*. The other bands are associated with Phi297 and D3112 Pa Phage predominantly. It is therefore clear that the dot matrix is mainly scaled to the high number of F10-like phages presented in this study.

The comparison of the sequences using Mauve shown in Figures 4.7 – 15 displays that the phages omit region specific to certain genes most are hypothetical and thus raises the question on the functionality of those genes. Even though we identify 35 F10 like phages it shows that the online database is not rich enough to compare fairly similar phages that still display large variation within their genomes. Based on the Mauve alignment comparison it could be inferred that all the phages displayed are novel phages

and would aid the NCBI database when uploaded once further curation of the phages has been carried out. The k-mer graphs as mention above allowed prediction of the number of phages however; due to low sequence reads possibly related to the titre of these phages the assemblers did not assemble complete phages for all the peaks. We can however gauge that the k-mer graphs do not fully represent the total number of phages present. Sample 36 presents in the k-mer graphs as one peak but when assembled we find 2 phages D3112 and Phi297 at roughly the same mean coverage (284 and 207 respectively).

Based on the aims of this chapter it can be concluded that the meta-virome assembly of phages genomically similar, is difficult and requires further analysis of sequence coverage and sequence similarity to be confident in the assemblies. The k-mer based method of removing low reads lengths and separating k-mers based on count and abundance helps assembly of genetically different phages but not phages sharing conserved regions. Through this we would add novel phages as draft genomes to the nucleotide database adding value to future blastn searches for Pa temperate phages.

5. Characterising the interactions of *Burkholderia* phage and isolating their genomic sequences for genome assembly.

5.1 Introduction

5.1.1 *Burkholderia* background in Cystic Fibrosis

Burkholderia cepacia was first described by Walter Burkholder of Cornell University in 1949 when he determined it to be the cause of bacterial rot of onion bulbs. The bacterium was originally named *Pseudomonas cepacia* and later changed to its current name (discussed in detail in chapter 1.3). *Burkholderia* are rod-shaped, free-living, motile Gram-negative bacteria ranging from 1.6-3.2 μm in length.

They possess multitrichous polar flagella as well as pili used for attachment. *Burkholderia cepacia* can be found in soil, water, and infected plants, animals, and humans (discussed in detail in chapter 1.3) (Miller et al, 2002).

Burkholderia cepacia complexes (Bcc) are key bacteria associated with microbial infection of the Cystic Fibrosis (CF) lung. The word complex has many paradigms and as Mahenthiralingam *et al.* (2002) states in their review '*B. cepacia'* are complex with key prominence on complex (Mahenthiralingam et al, 2002). This statement from the authors highlights the changing perception and understanding of this group of bacteria over the last 2 decades (Mahenthiralingam et al, 2002). It therefore also raises the question about how much do we understand about the bacteriophages that infect this bacterial species. Bcc do not infect non-cystic fibrosis patients readily and colonisation is not usually chronic. A case report of a mother of two CF individuals showed that she unfortunately became infected with the same epidemic strain of *B cepacia* acquired from one of her CF children (Ledson et al, 1998). This case report highlighted worrying developments for partners, siblings and parents of CF individuals who might be at risk of transmission of the pathogens carried in the lungs of CF patients (Ledson et al, 1998).

Burkholderia species is currently comprised of 20 sub species of which 10 species are commonly found in the lungs of CF patients (see chapter 1.5 for more detail)

(Coenye et al, 2001; Cystic fibrosis foundation., 2016). Bcc is more problematic to clear possibly due to increased antimicrobial resistance (Kuti et al, 2004). The *Burkholderia* genomovars have been linked to varying virulence, some study indicate *B. cenocepacia* being more virulent (Aris et al, 2001; De Soyza et al, 2001). Bcc genomovar III have been shown to replace *multivorans* species colonising in the lung CF patients (Mahenthiralingam et al, 2001). *Burkholderia* genomovar III has been shown to be the most prevalent subs species of *Burkholderia* isolated from CF patients, this can also vary geographically (Mahenthiralingam et al, 2002).

Bcc disease or Bcc syndrome has high mortality rates this is also similar to patients that are chronically infected with *B. multivorans* (Zahariadis et al, 2003). A case study of a patient showed that within week of acquiring Bcc along with Pa being present the forced expiratory volume (FEV) dropped from 43 % to 22% despite being on multiple antibiotics (Shafiq et al, 2011). Another case study showed that after a boy initially became colonized with *B. multivorans* at the age of 7 and developed cepacia syndrome when he was 16 years old (Blackburn et al, 2004). This case study also highlights that cepacia syndrome is not confined to cepacia genomovars and that the syndrome occurred while antibiotics were being administered to this particular patient (Blackburn et al, 2004). Antibiotic resistance is a major contributor to Bcc virulence and failure to manage infection with strains being resistant to β-lactams antibiotics (Trepanier et al, 1997). In up to 17% of clinical cases β-lactam based antibiotic fail to treat melioidosis caused by *B. pseudomallei (Chierakul et al, 2005)*.

5.1.2 *Burkholderia* genome and their phage

The *Burkholderia* genome consists of multiple replicons (chromosomes) which can vary between strains. The *Burkholderia cepacia* type- strain ATCC 25416 (genomovar I) is 8.1 Mb in length and is known to have four circular replicons (rrn). Its largest replicon contains 4 rrn operons and the other two megabase- sized replicons contain a single rrn operon each. From this information it can be inferred that the

organism has three chromosomes and one large plasmid (Rodley et al, 1995). Genome Sequence of *Burkholderia cenocepacia* H111 consists of 3 chromosomes (Carlier et al, 2014), *Burkholderia cenocepacia* J2315 consists of 3 chromosomes and a plasmid (Holden et al, 2009) as does *Burkholderia multivorans* ATCC 17616 (Nishiyama et al, 2010) and *Burkholderia vietnamiensis* G4 consists of 3 chromosomes and 5 plasmids (Maida et al, 2014). This also means that each chromosome can have its own lineage. One study showed that within the *Burkholderia* species Bcc strains showed the largest sequence homology 78 to 63 % of its coding sequences (Holden et al, 2009). A study investigating the degree of sequence homology showed it was greatest on the largest chromosome with the other chromosomes exhibited more divergence (Holden et al, 2004).

The implications of this can be that a temperate phage would integrate into a specific chromosome. *B. pseudomallei* K96243 is a 7.3 Mb which appears to have three prophage regions, where only one (ΦK96243) is known to be fully functional as a bacteriophage (Holden et al, 2009). *B. pseudomallei* E12 has shown to spontaneously produce two temperate bacteriophage ϕE12-1 and ϕE12-2 (Ronning et al, 2010). Therefore, it can be hypothesised that some *Burkholderia* harbor multiple inducible prophages. Bacteriophages in *Burkholdria* species can act as a means of transposable elements and carry moron genes that may aid their host (Ronning et al, 2010). Bacteriophage 1710b-3 has been predicted as carrying a moron gene that has abortive infection function (Ronning et al, 2010). Burkholderia phage 17616-4 has predicted moron genes that may be involved in pyocin-related function (Ronning et al, 2010).

BcepMu is a prophage present in *Burkholderia cenocepacia*, which is a Mu-like phage (Summer et al, 2004). The BcepMu phage replication and packaging method results in an additional 2 kbp of the host bacterial DNA in the virion head (Summer et al, 2004). This hypothetically increases the potential level of transduction and transfer of genetic material as the phage also integrates almost randomly into the host cell (Summer

et al, 2004). These phages could potentially transfer any region of the *Burkholderia* genome sequence to another host cell.

Aims: In this study 47 CF samples of *Burkholdria* cepacia complex have been obtained from the Freeman hospital to induce temperate phages using norfloxacin, characterise and genome sequence the inducible Bcc phage. This study aims were to show the cross infection profiles and to characterise the genome of the Bcc temperate phages. The study aims to use the k-mer peak separation method and construct draft genome to submit to the online nucleotide database. This study shows the chemical induction of temperate phages from 47 Bcc clonal isolates equating to 6 of the 10 sub species of Bcc (Table 2.2.3) mainly consisting of genomovar III A (n=22) and multivorans (n=13).

5.2 Results

5.2.1 Cross Infection of phage lysate onto 47 Bcc backgrounds

Bcc phage lysate were prepared using the same protocol as described in chapter 2.3.1. Phage lysates from 47 CF clinical Bcc isolates were spotted onto all 47 of the available bacterial backgrounds including the originating complex background from which they were induced. Due to insufficient clinical data deeper characterisation as presented in chapter 3 of this study was not possible. The percentage mean infection of each lysate containing Bcc phages from each of the complexes were compared and no significant difference was seen, thus this data is not presented for each sub species. The Bcc samples as mentioned in the introduction (chapter 5.1.2) consist of 22 genomovar III A and 13 multivorans. This constitutes 35 of the 47 samples, from this we can infer that genomovar III A is the most prevalent sub species in the 47 CF patients presented in this study, reflecting current findings based on prevalence.

The mean infection of the phages lysates induced form the genomovar III A is 64 % and for multivorans is 65% across the 47 Bcc bacterial cohorts. The results were tabulated as infectivity (red, yellow and green) or no infectivity (black) in Figure 5.1. The key information that we can take from the cross infection table is that all the Bcc bacterial isolates harboured at least one inducible temperate phage. The phage induced from Bcc 16 exhibited the highest infection rate (96%) and was a phage from a *Burkholderia vietnamiensis* bacterial isolate. It is similar to the KS9 phage and a genome alignment is shown in figure 5.10. The phage from sample 20 exhibited the lowest infection rate of 38%. The sequence sample of this lysate showed 2 contigs at 27 and 26 kbp with no sequence similarities on the Viruses database and was not presented in this study as they were considered as partial phage genomes. The Bcc bacterial isolate that was most susceptible to infection (98%) was Bcc 17 (*B. multivorans*) and the highest incidence of clear lysis due to phage infection was seen on Bcc 34 (*B. multivorans*). Figure 5.1 also shows that 68% of the Bcc phages could re-infect their originating host.

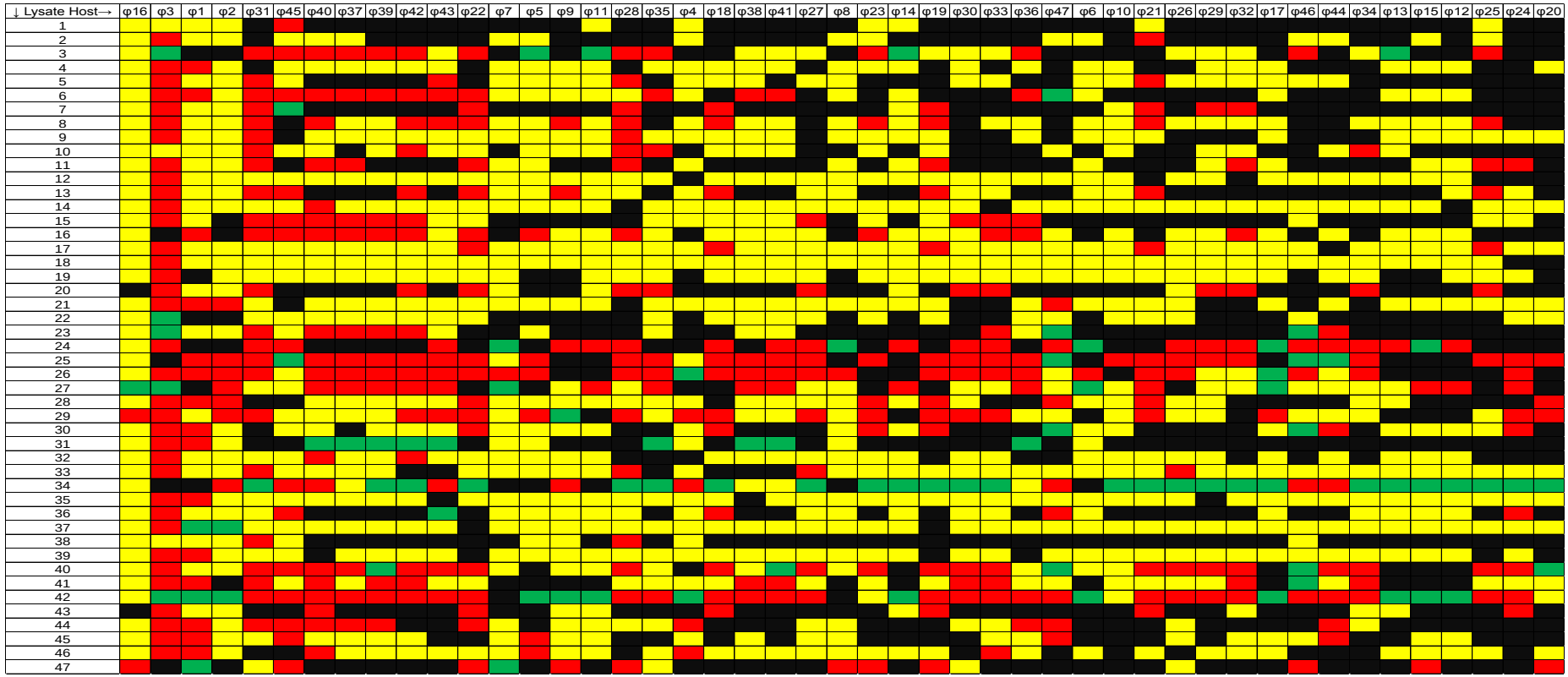
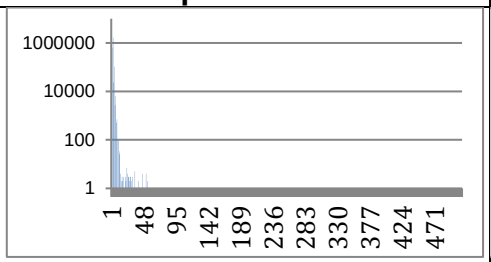
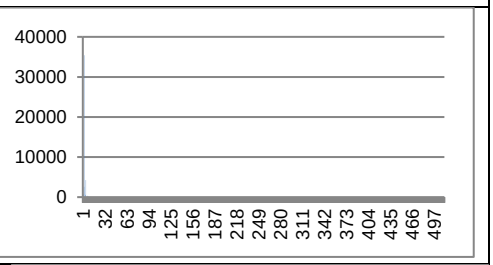
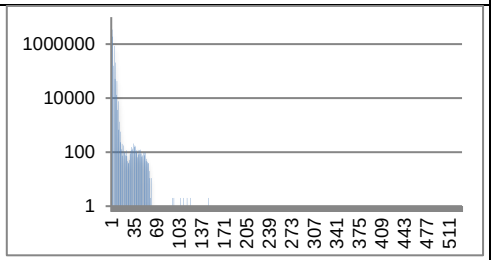
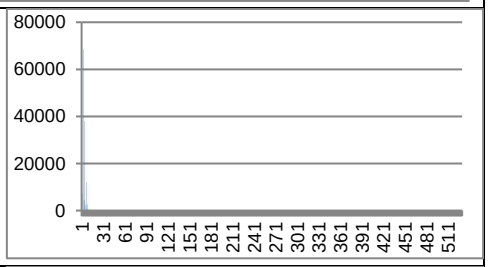
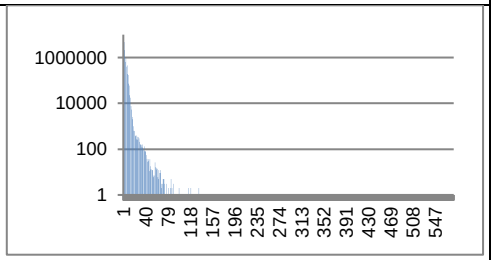
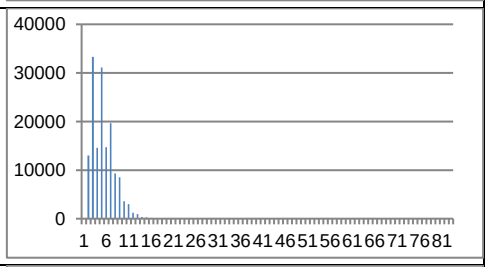
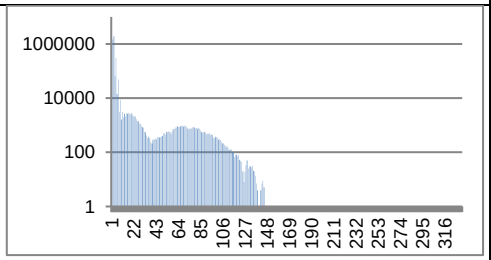
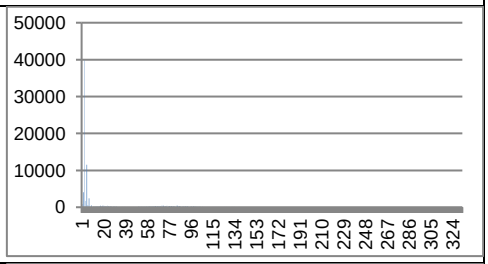
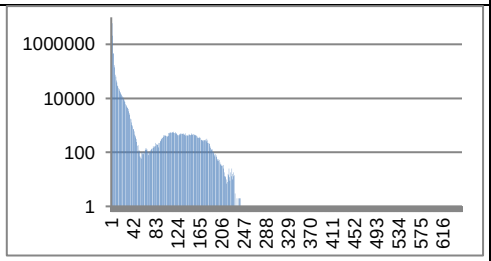
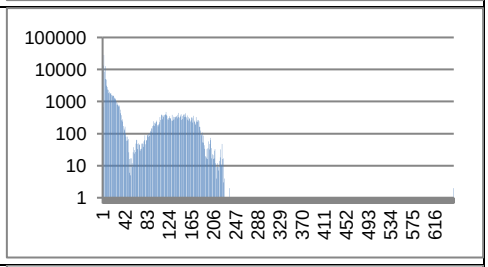
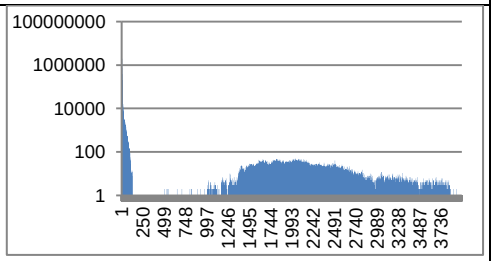
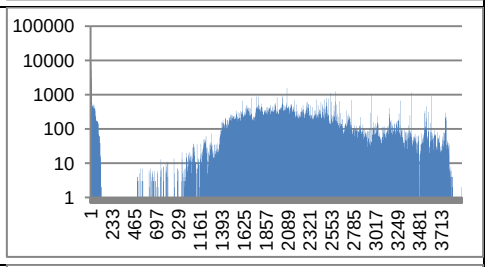
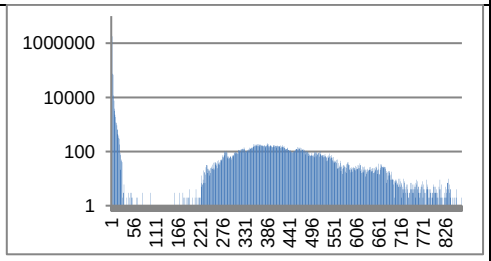
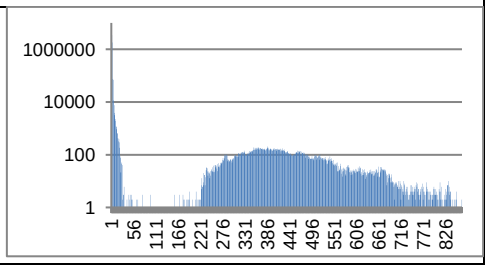


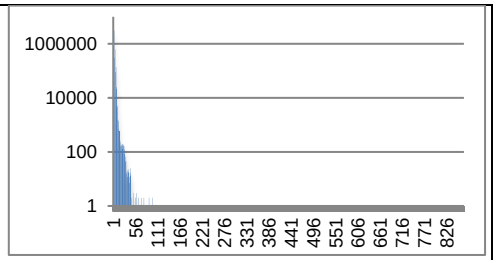
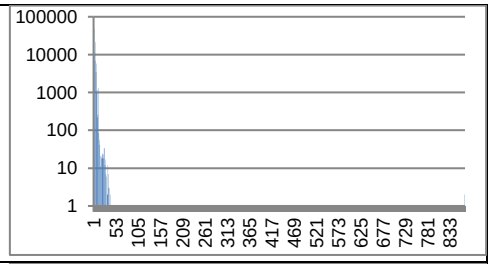
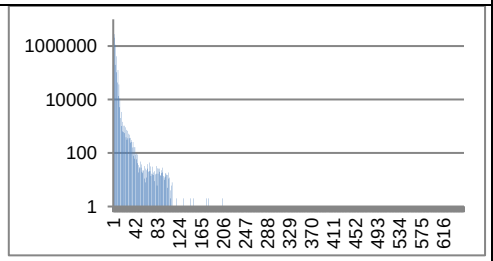
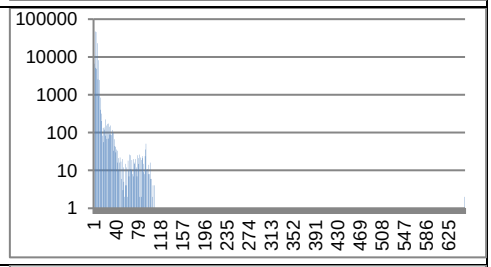
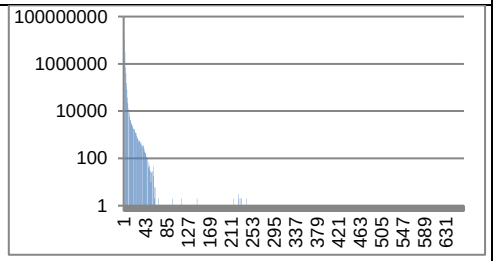
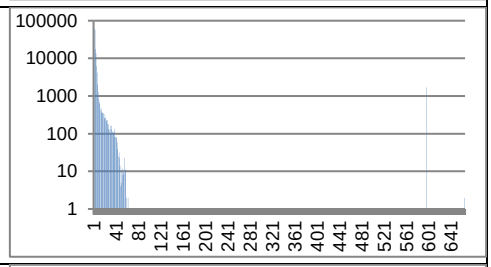
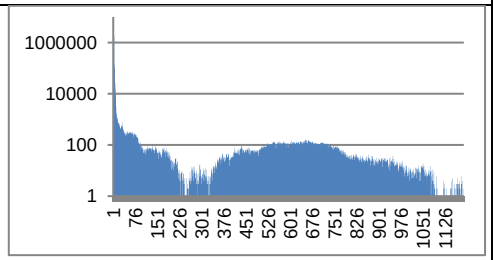
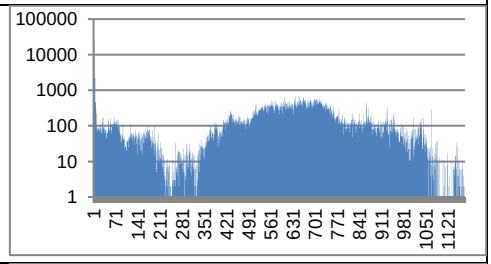
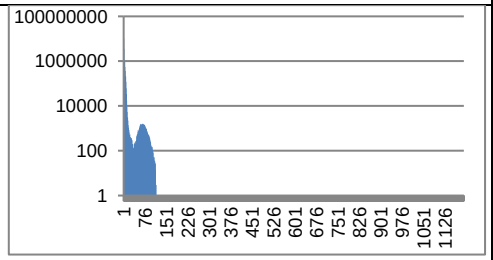
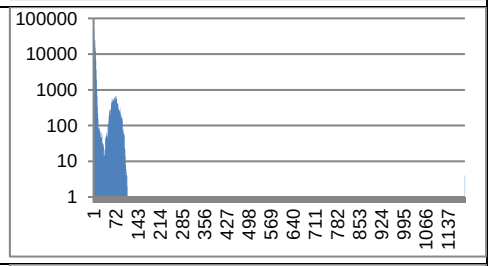
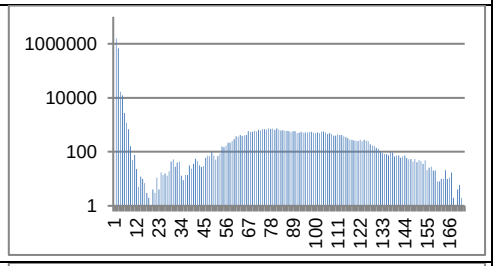
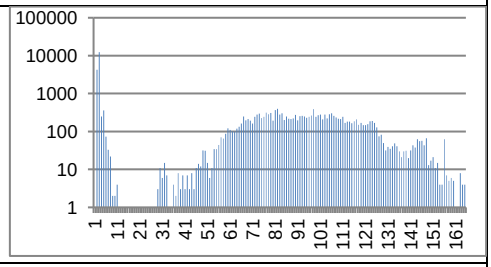
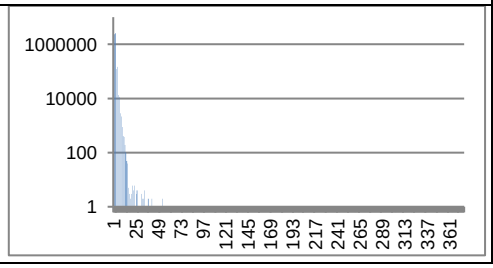
Figure 5.1: Infection profile of Bcc phage. Schematic diagram showing cross infection assay of induced phage lysates onto 47 Bcc bacterial lawns. The table shows each induced phage lysate was spot tested onto all 47 Bcc isolates. The Yellow (turbid), Red (lysis) and Green (halo) all represent Phage infection and Black denote the inability of the phage to infect the bacterial background. The phages are shown based on infectivity highest to lowest left to right.

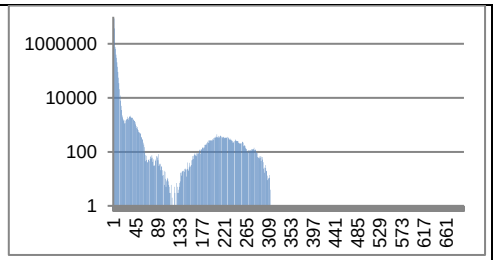
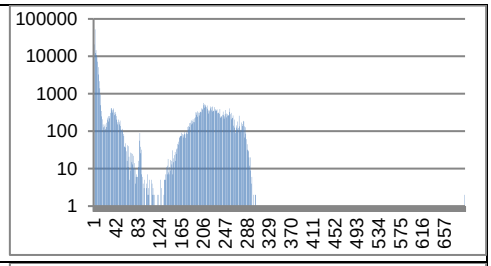
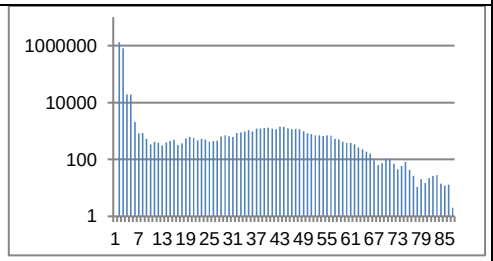
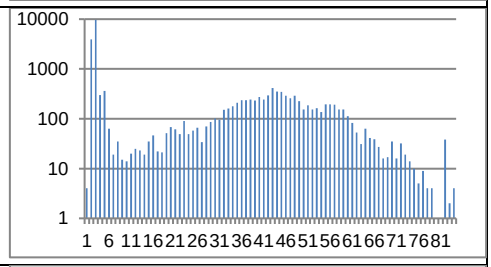
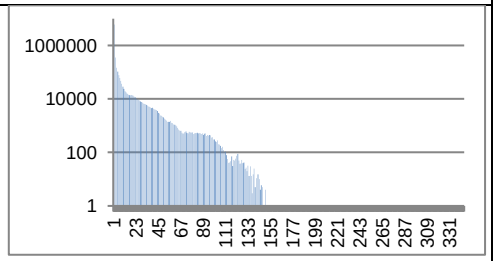
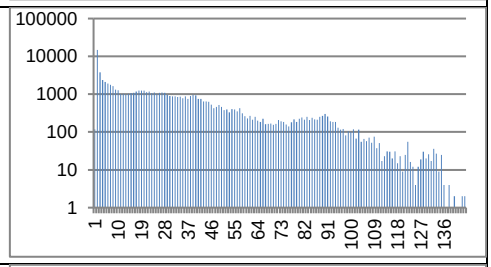
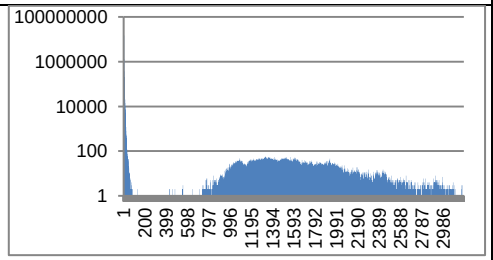
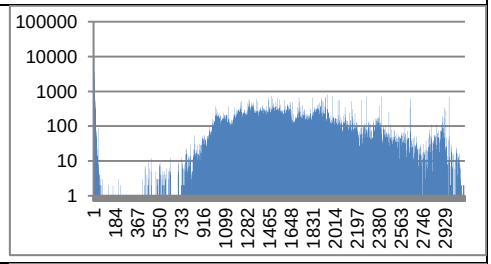
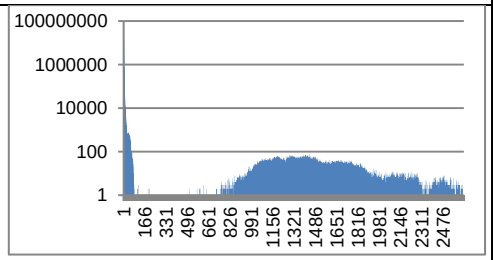
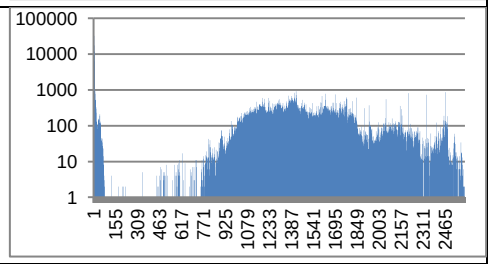
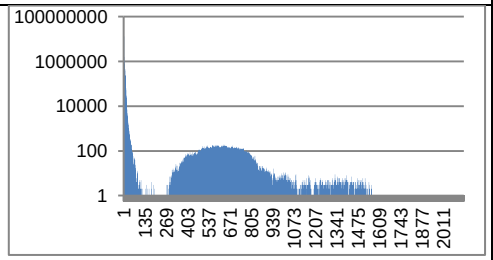
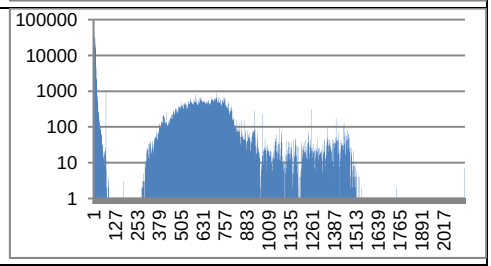
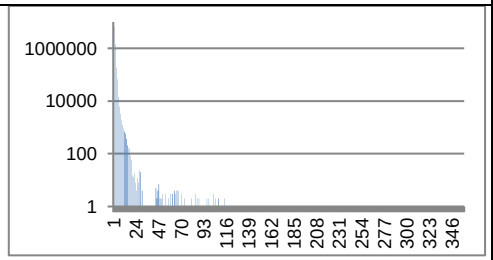
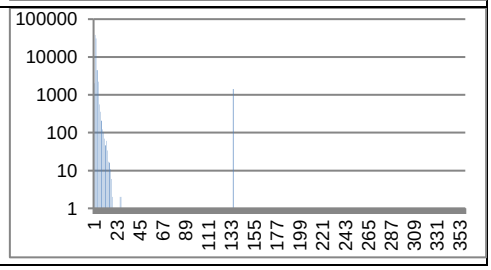
5.2.2 Genome characterisation

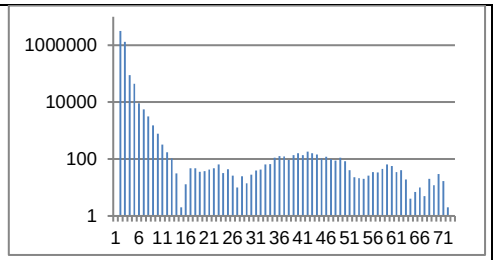
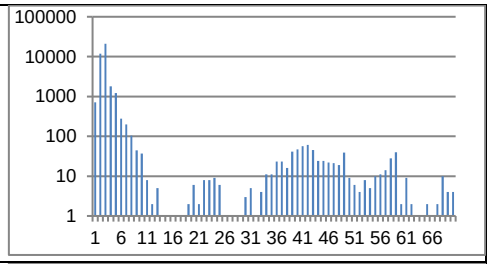
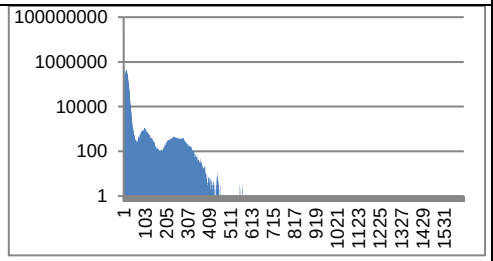
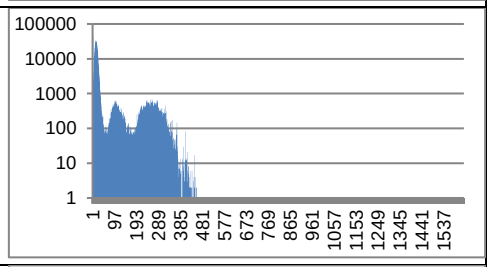
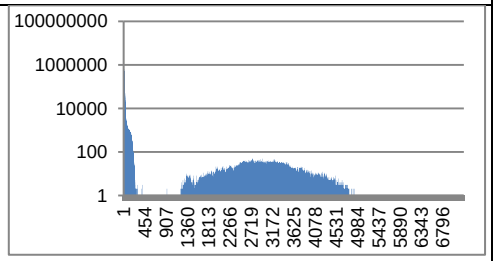
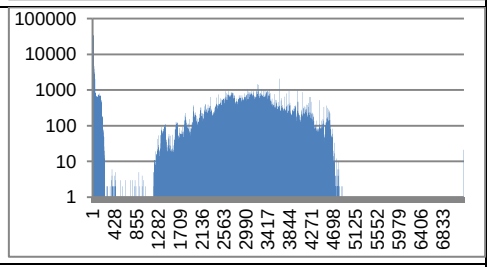
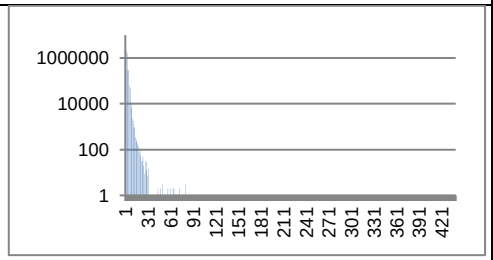
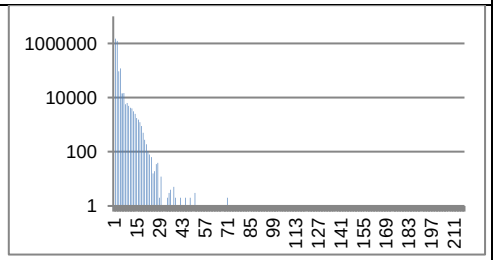
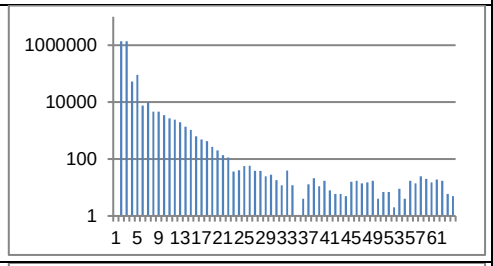
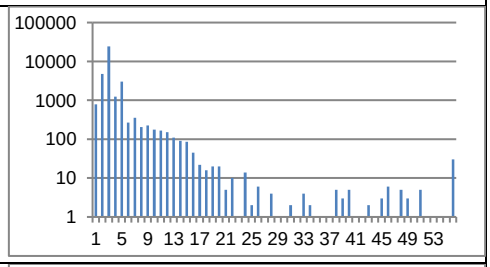
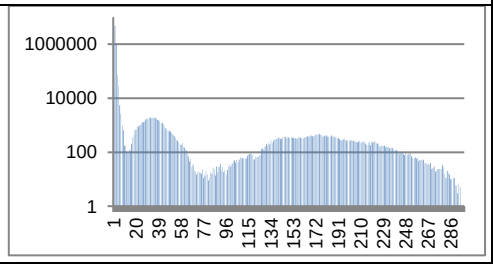
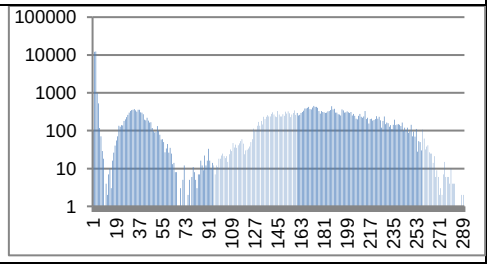
Phage genomic DNA was extracted from the Bcc phage lysate samples using the method described in 2.6.2 and sequenced as described in 2.6.3. The low level erroneous k-mers and bacterial contamination was removed using the khmer toolkit (chapter 2.6.5) and the k-mer graphs are shown in table 5.1. Using the extracted k-mers the genomes were assembled using SPAdes v3.5.0 (chapter 2.7). The assembled genomes were searched for sequence similarity using Basic Local Alignment Search Tool nucleotide (BLASTn) virus database standalone command line BLAST against the virus database (chapter 2.11.1). In this format the sequence identity could be focused specifically on viral genomes independent from the non-redundant database at NCBI. Searching locally in this manner using blastn significantly reduced the time taken to generate comparison output files.

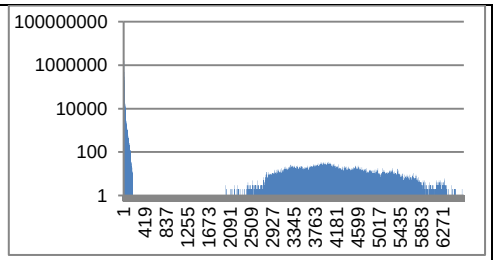
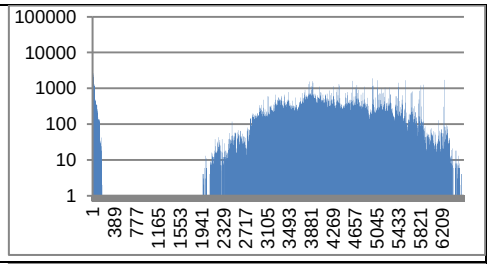
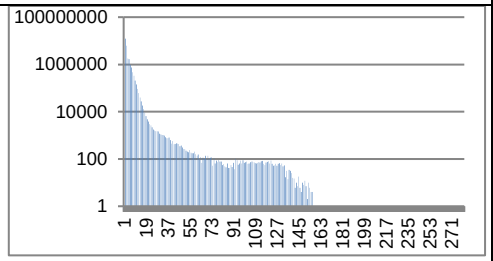
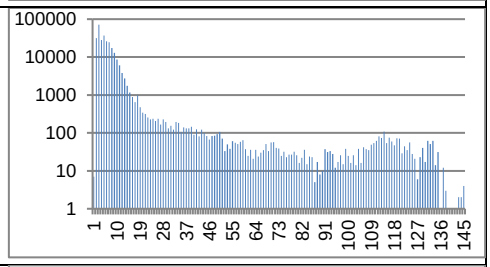
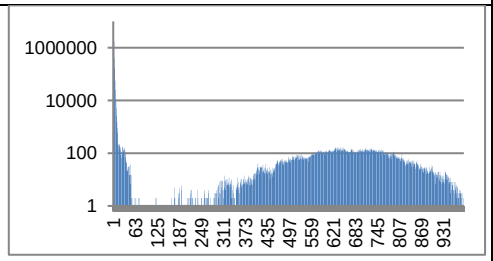
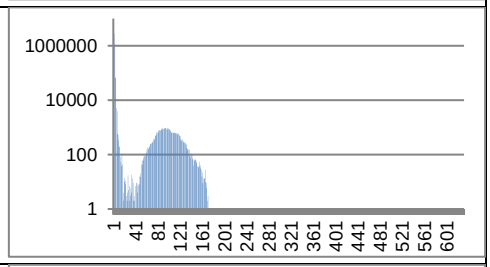
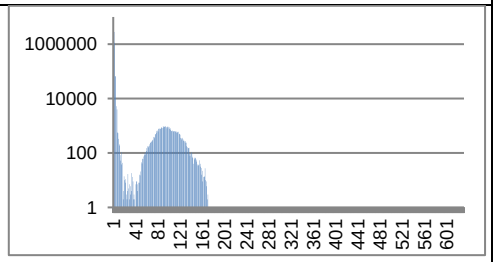
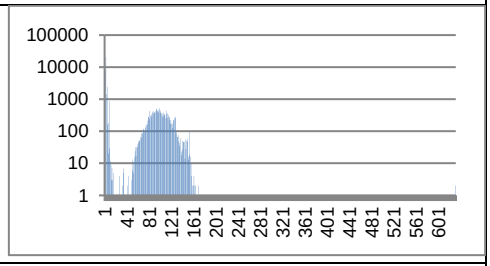
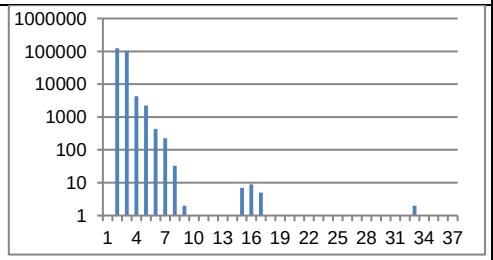
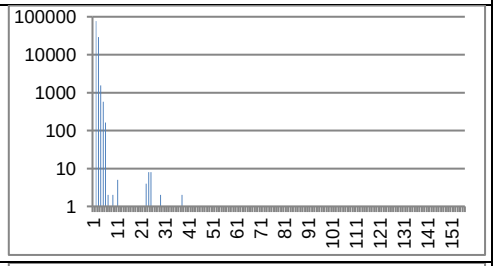
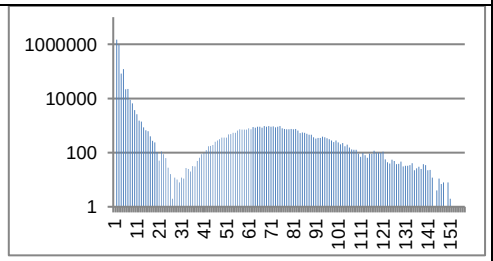
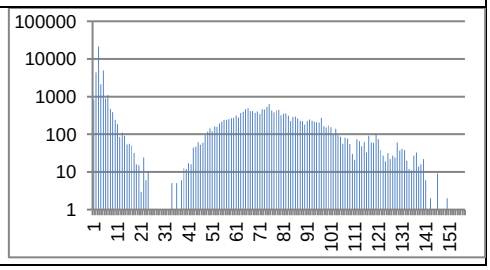
The blastn similarities were used to group and compare the phages found in the Bcc sequence data using a dot plot generated in gepard and mauve alignment. Figure 5.2 shows the sequence comparison in a dot matrix using gepard comparison software. Each phage has been concatenated end on end and compared. Therefore, the diagonal line through the centre shows is solid as a direct comparison of phage versus phage and where there are lighter areas the lower the homology between the genome sequences. Table 5.2 shows the number of phages that shared a sequence similarity using blastn.

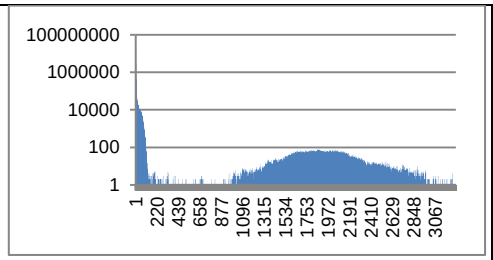
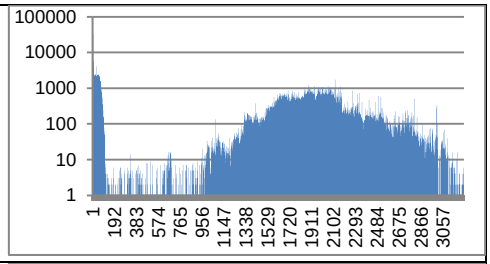
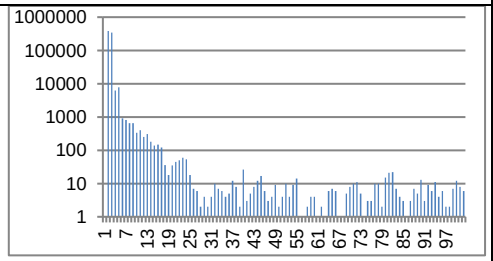
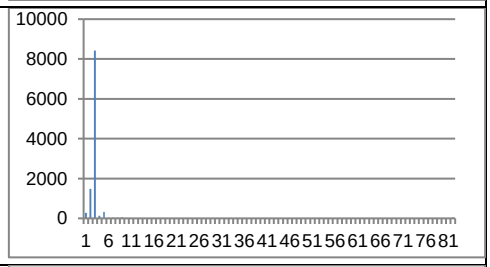
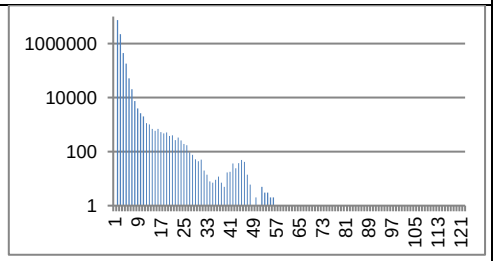
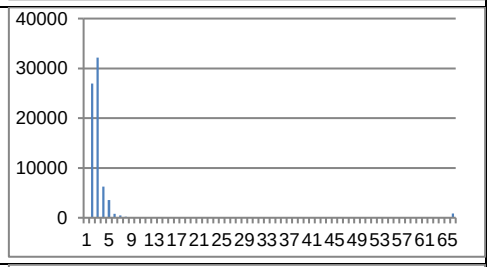
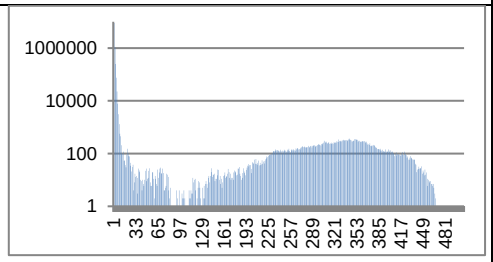
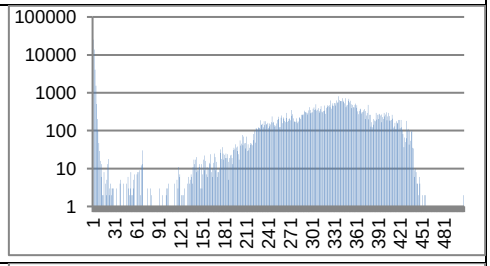
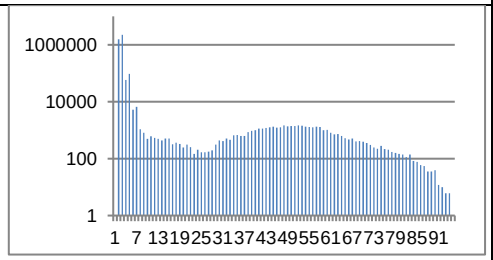
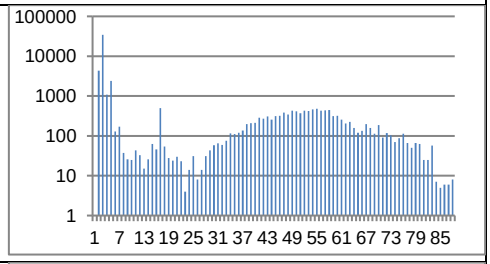
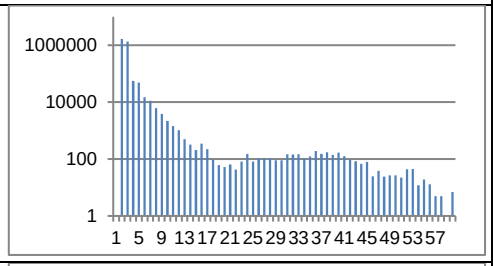
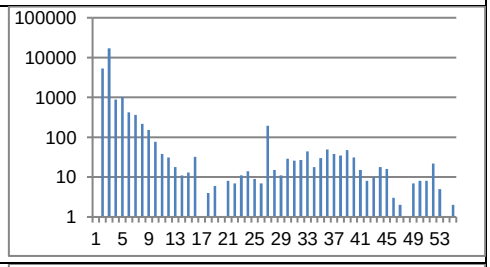
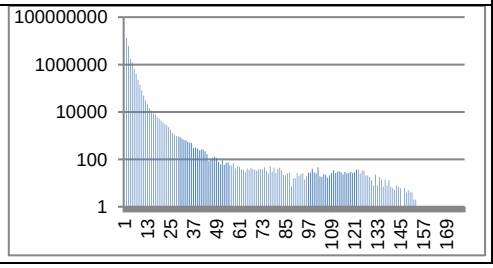
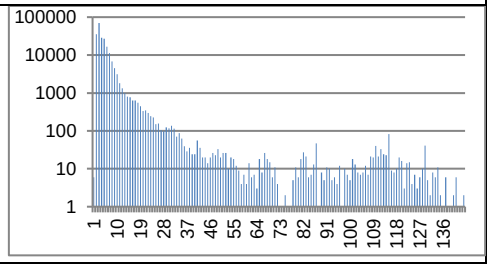
Sample number	Log scale of k-mer graph pre error peak removal	k-mer graph post error peak removal
1		
2		
3		
4		
5		
6		
7		

8		
9		
10		
11		
12		
13		
14		Not enough sequence data.

15		
16		
17		
18		
19		
20		
21		

22		
23		
24		
25		Not enough sequence data.
26		Not enough sequence data.
27		
28		

29		
30		
31		
32		
33		Not enough sequence data.
34		Not enough sequence data.
35		

36		
37		
38		
39		
40		
41		
42		

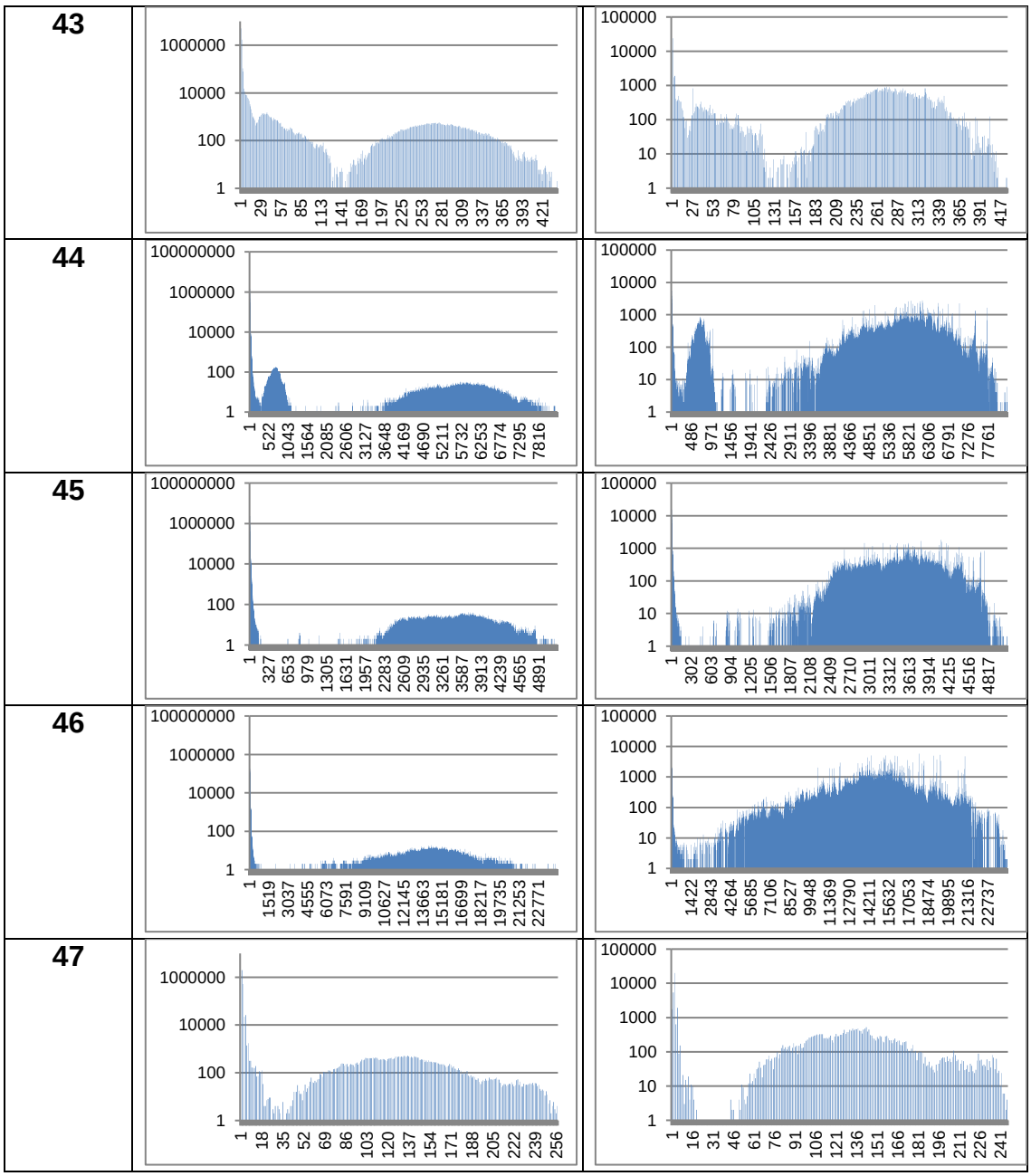


Table 5.1: Bcc K-mer peaks pre and post sequence extraction. The table shows the pre and post k-mer graphs. The post sequence extraction graphs illustrate k-mers after removal of low abundance k-mers

This study identified seven phages which were shown to share a degree of sequence homology with the phage Bcep176. The level of nucleotide homology between the four different phage genomes similar to Bcep176 can be seen in figure 5.3. The highest sequence homology is seen by vB_Bcc_11a and the least by vB_Bcc_5a. Both vB_Bcc_11a and vB_Bcc_5a are similar in sequence length to Bcep176. The other two

phages vB_Bcc_23a and vB_Bcc_35a are shorter by 4 and 6 kbp in length respectively. The purple block seen in vB_Bcc_11a and vB_Bcc_23a shares some sequence homology to the reference sequence Bcep176. At around the 31 kbp region of the Bcep176 phage there is a gene (identified as hypothetical gene) which is not present in vB_Bcc_11a (20 kbp region) and vB_Bcc_23a (14 kbp region).

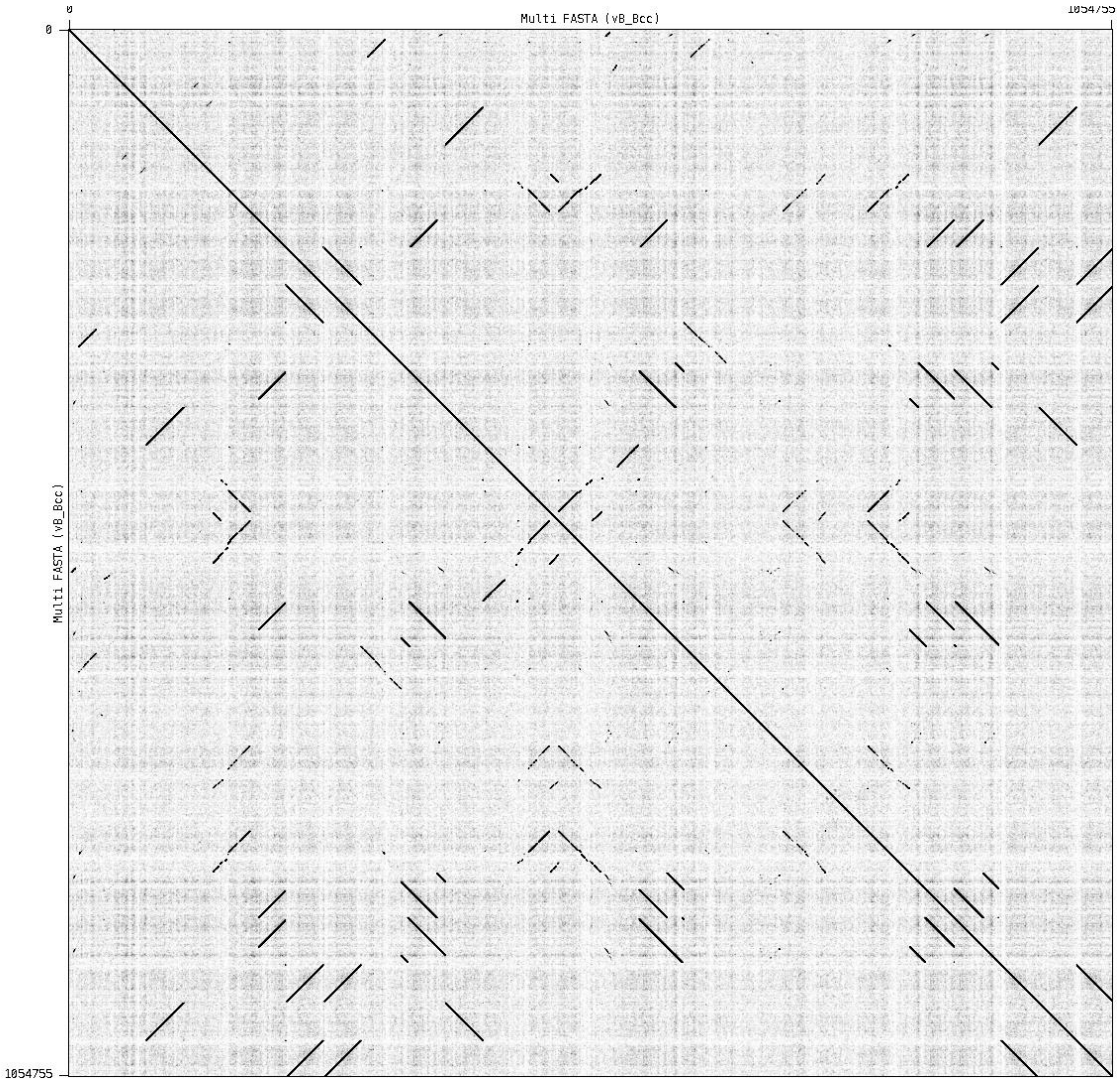


Figure 5.2: **A dot plot of 26 Bcc phages.** A dot matrices comparing the nucleotide sequences of the 26 Bcc phages. The x and y-axis consists of 26 phages that were concatenated together into a fasta file. This map was generated using Gepard v1.40 (Krumsiek et al., 2007) using a window size of 25 and word length of 11

Most similar phage on blast database (family)	Genome size kbp	Blastn nucleotide match in kbp for the largest region of similarity	Number of phages similar to the phage
JG068 (*Podoviridae*)	41.604	0.790	1
KS9 (*Siphoviridae*)	39.805	0.786, 0.788 and 11.771	3
Bcep176 (*Siphoviridae*)	44.856	0.127, 10.374, 8.265 and 0.699	7
BcepMu (*Myoviridae*)	36.748	36.747	4
KS10 (*Myoviridae*)	37.635	37.623, 37.632 and 37.627	3
KS5 (*Myoviridae*)	37.236	8.406	1
PhiE12-2 (*Myoviridae*)	36.690	2.606	1
J2315 ch2	3200	35.805	1
DC1 (*Podoviridae*)	61.847	0.307	1
PhiRSA1 (*Myoviridae*)	38.760	1.048	1
Salicoliphage CGphi29/ Bcep781 (*Myoviridae*)	40.695/48.247	0.127/0.043	1
PhiE125 (*Siphoviridae*)	53.166	0.751	1
Phi1026b (*Siphoviridae*)	54.845	0.770	1

Table 5.2: Bcc phage similarity. The table shows the blastn similarities on the viruses' database. The total number of Bcc phages is 26; the table shows the variation in blastn hits to the Bcc phages.

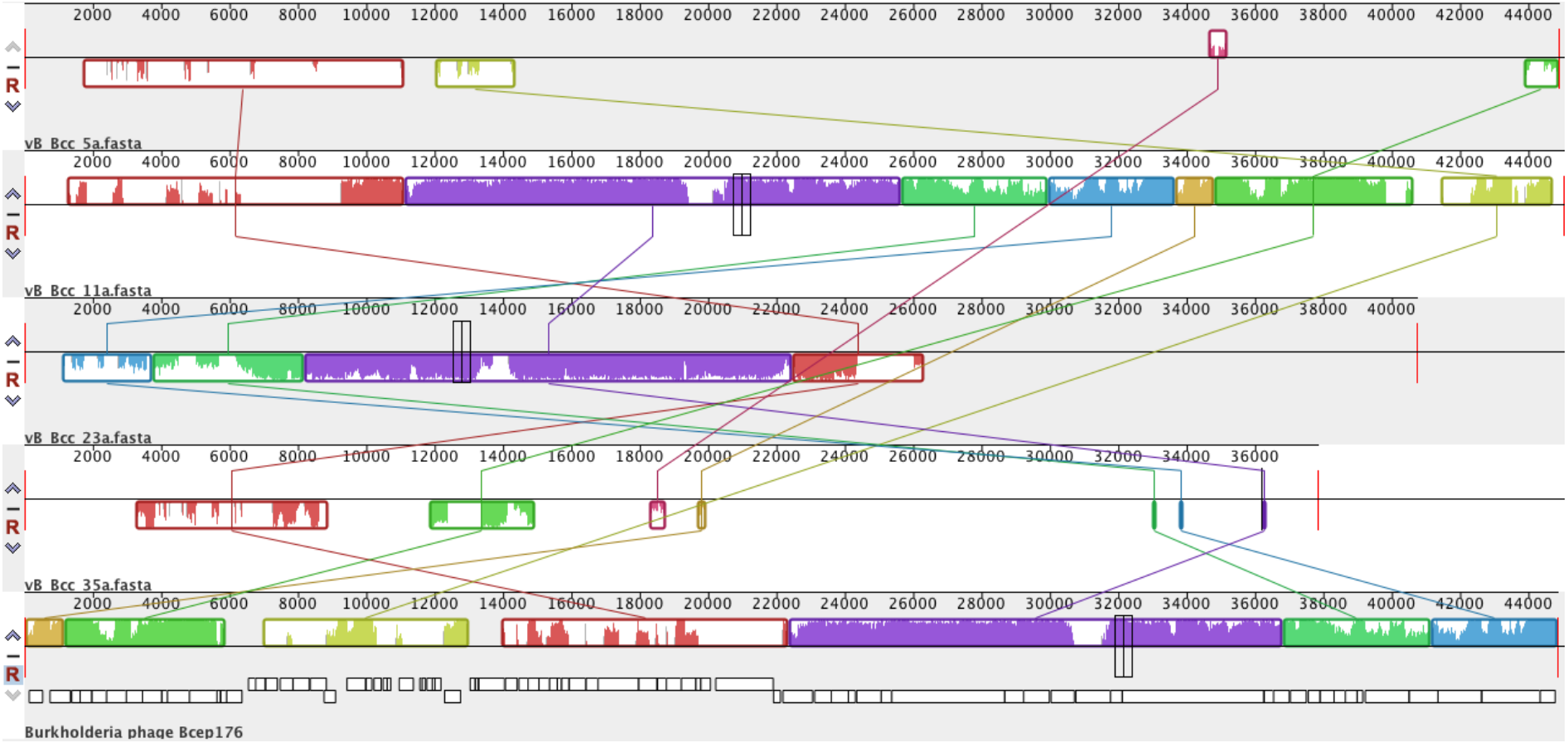


Figure 5.3 Mauve alignment of Bcep176-like phage. A sequence alignment that compares the nucleotide sequence of 4 Bcep176–like (vB_Bcc_5a, vB_Bcc_11a, vB_Bcc_23a and vB_Bcc_23a) *Burkholderia* phages with Bcep176. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

Figure 5.4 shows an example of one of the 4 phages that shared almost 100% sequence homology to the phage BcepMu. However, the genome of this phage was larger than the reference BcepMu phage. The sequence of the isolated phage is larger at the left end of the phage by around 4 Kbp. There are 3 extra genes identified on the left hand of the genome and one gene on the right region of the genome. The first putative gene is identified as a hypothetical protein, the second is identified as an integrase core domain protein and the third putative gene is identified as a small-conductance mechanosensitive channel. The extra putative gene identified at the right end is identified as a squalene-hopene cyclase. This is indicative of BcepMu phage DNA packaging where fragments of the genomic DNA are incorporated into its phage capsid and has been previously reported (Summer et al, 2004).

The Bcc phage vB_Bcc_43a shared very little sequence homology to any phage on the viruses' database, the Salicoliphage CGphi29 shared a 127 bp similarity and only 43 bp sequence homology shared to Bcep781. Figure 5.5 shows the mauve alignment of these two phages to vB_Bcc_43a. The vB_Bcc_43a phage is substantially larger in length and may be a novel phage. The second reference sequence is used to support that this is potentially a *Bukholderia* like phage. The Bcep781 phage shares sequence homology to vB_Bcc_43a across a putative Holliday junction resolvase. Resolvases are used by phage for DNA recombination.

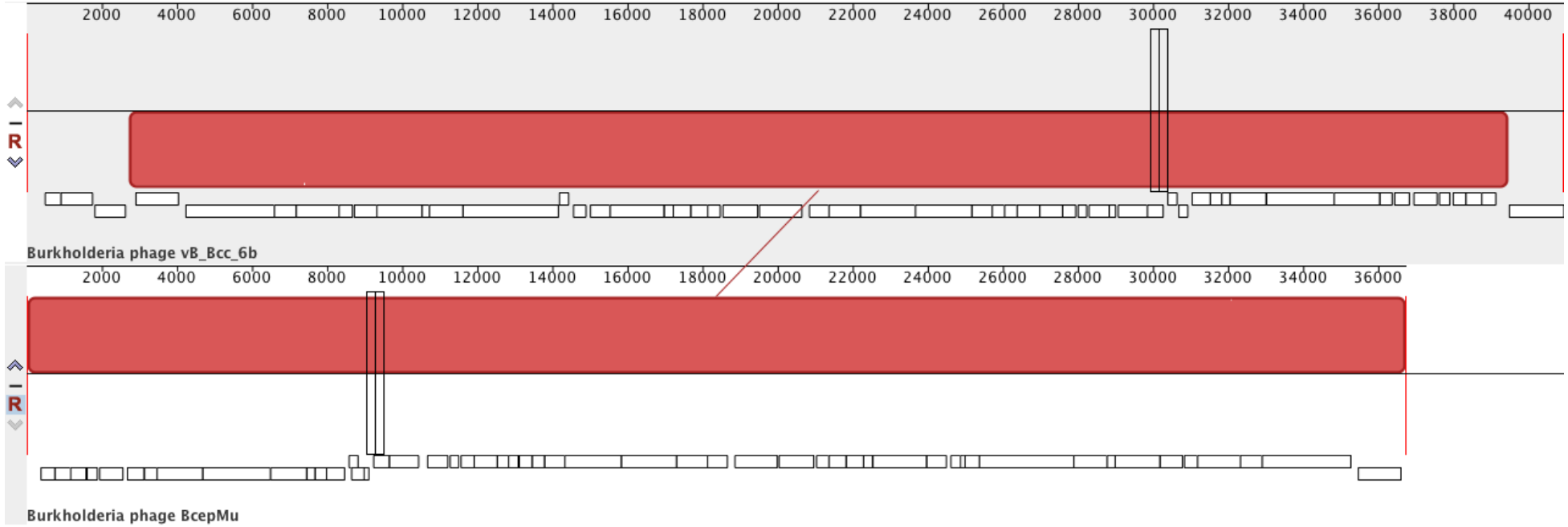


Figure 5.4 Mauve alignment of BcepMu-like phage A sequence alignment that compares the nucleotide sequence of a BcepMu–like (vB_Bcc_6b) *Burkholderia* phage with BcepMu. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

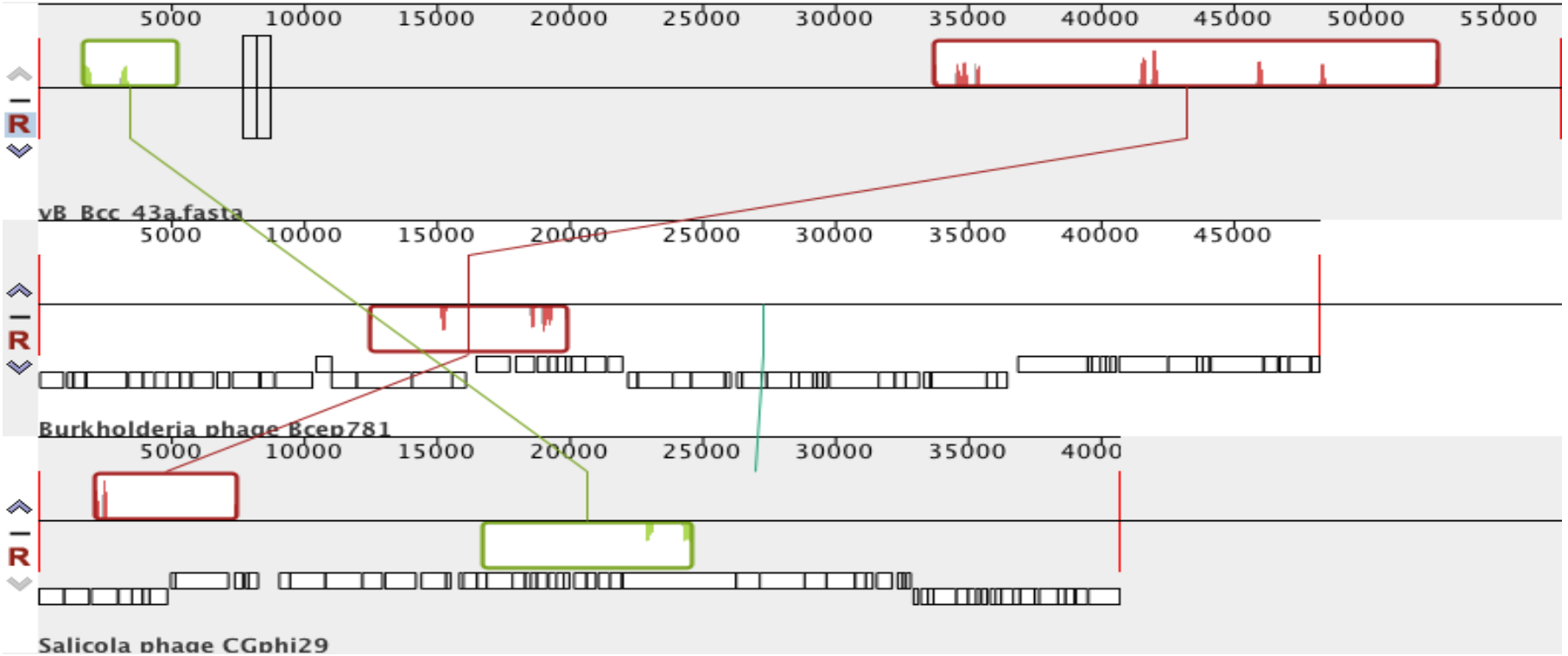


Figure 5.5 Mauve alignment of CGphi29-like phage A sequence alignment that compares the nucleotide sequence of a *Burkholderia* phages (vB_Bcc_43a) with Bcep781 and Salicola CGphi29. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

Figure 5.6 shows the sequence alignment of a DC-1 like phage, vB_Bcc_29a is a 35 kbp phage and DC-1 is over 60 kbp in length. They only share 307 bp sequence similarity based on the blastn search. The shared homology at the 60 kbp region on the reference DC-1 genome is across a putative lysozyme (99%). Lysozymes are used by phage to degrade the bacterial cell wall and facilitates phage infection.

The phage vB_Bcc_17a shares a sequence 38.805 kbp homology to a region on chromosome 2 of the J2315 *Burkholderia* genome. The viruses' database showed no similarities and therefore a remote alignment was carried out to show it was an inducible prophage. The mauve alignment is shown in figure 5.7.

The sequence homology between JG068 and vB_Bcc_4a is shown in the mauve alignment in figure 5.8. The shared blastn similarity is shown to be very limited with only 790 bp of similarity. The mauve alignment shows some homology in three distinct regions where the red block represents the largest region resemblance is seen. The sequence homology to the reference genome is seen on the 7.2 kbp region to a putative DNA helicase, 9.6 – 12 kbp DNA polymerase, 16 - 18.7 kbp RNA polymerase and the 19 – 26 kbp which shares homology to hypothetical head, head to tail connector, scaffolding protein, capsid protein and the beginning of the tail tubular protein. This homology is only evident in the Mauve alignment and not seen from blastn search results.

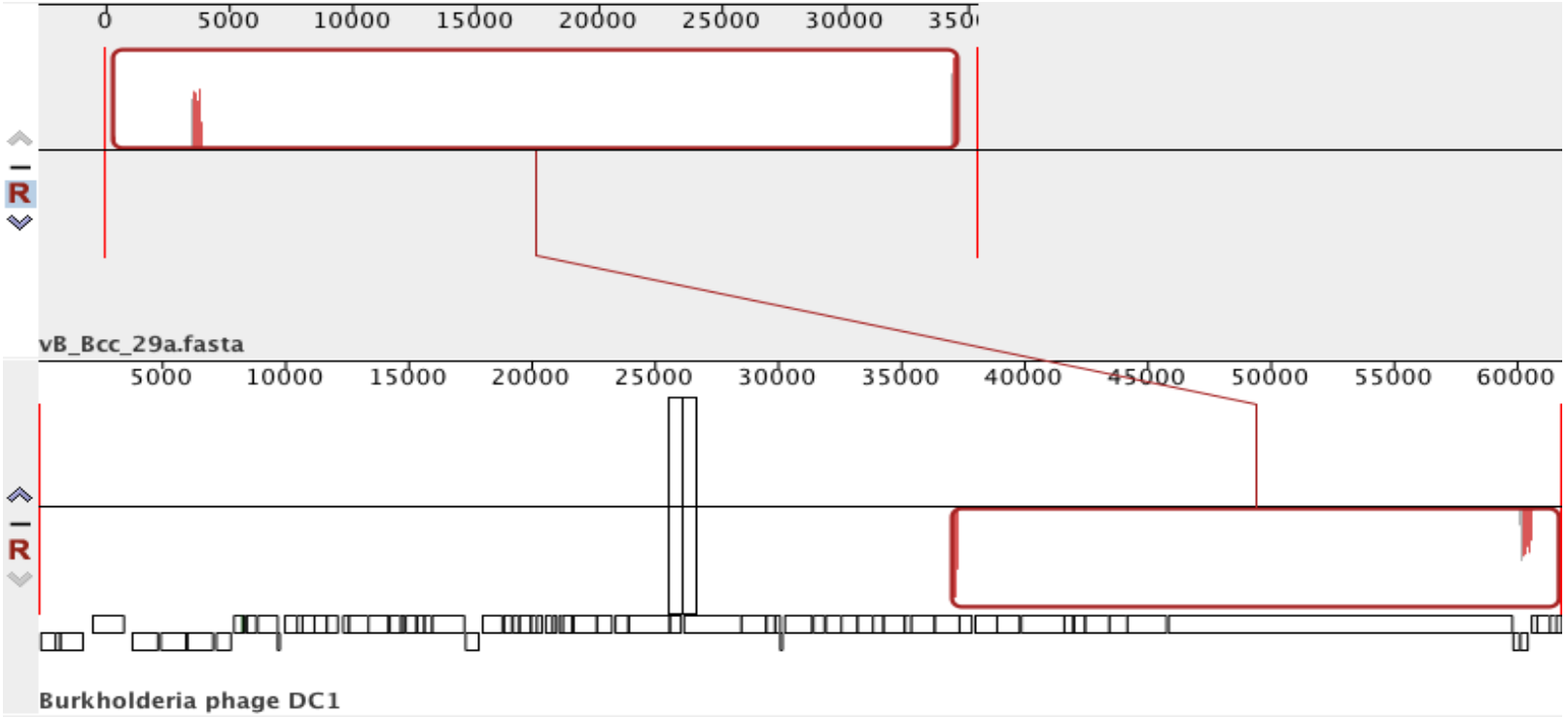


Figure 5.6 Mauve alignment of DC1-like phage. A sequence alignment that compares the nucleotide sequence of a DC1–like (vB_Bcc_29a) *Burkholderia* phages with DC-1. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

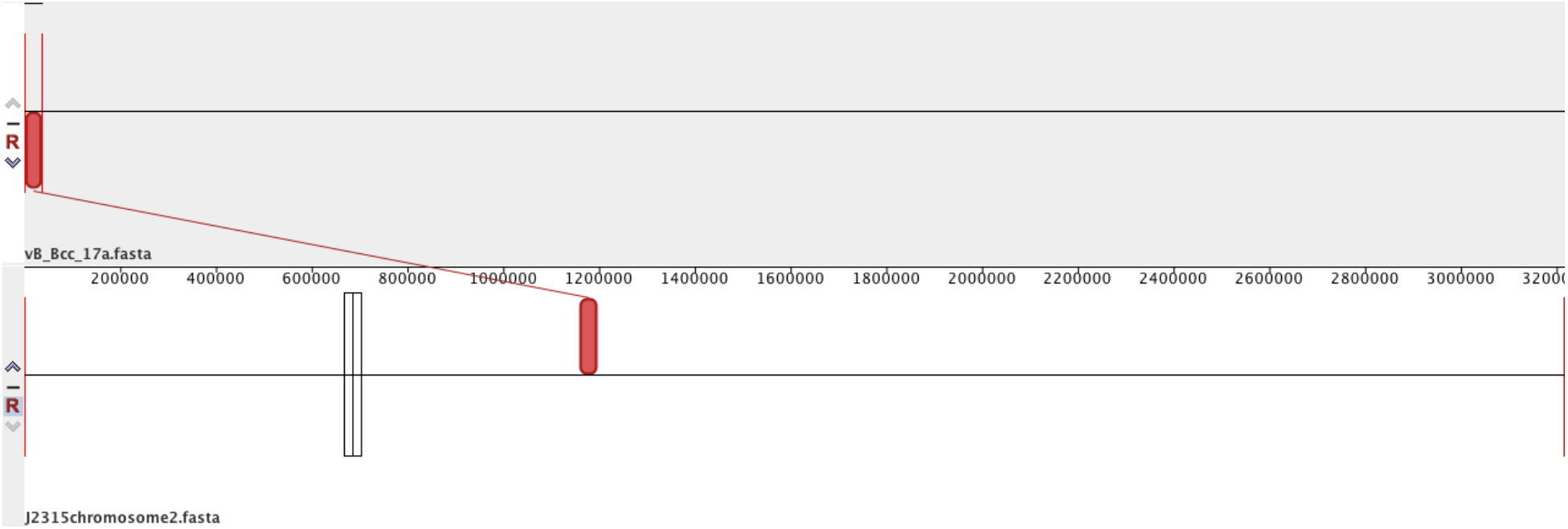


Figure 5.7 Mauve alignment of J2315ch2-like phage. A sequence alignment that compares the nucleotide sequence of a *Burkholderia* phages (vB_Bcc_17a) with J2315_ch2. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

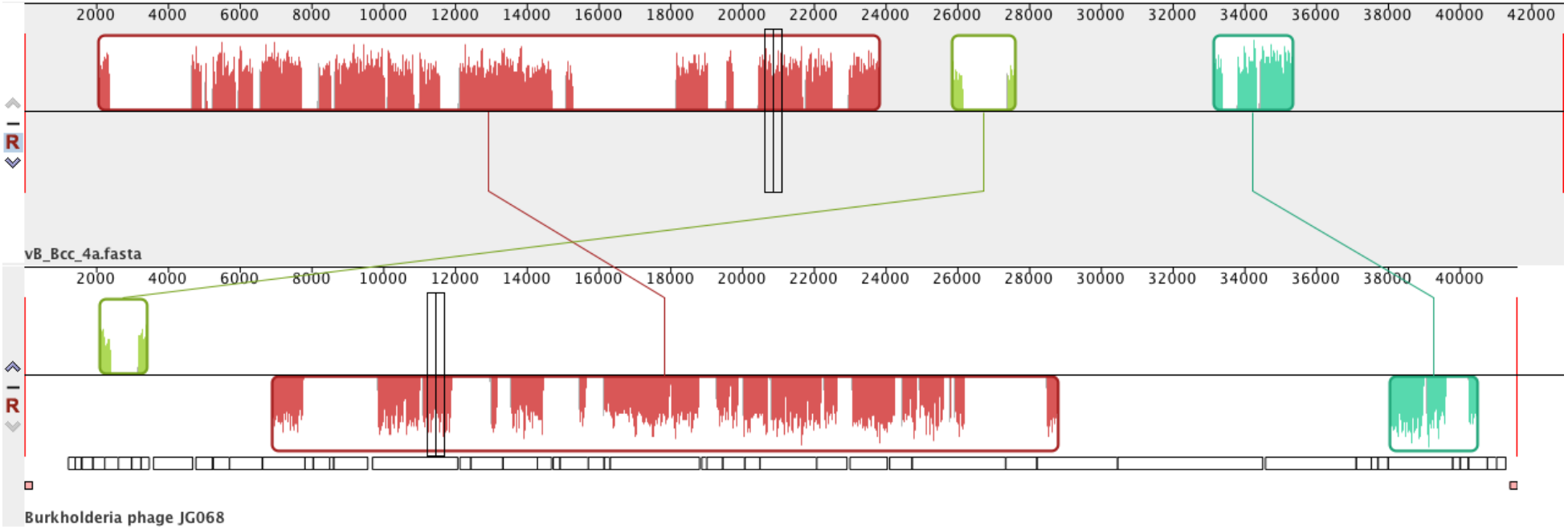


Figure 5.8 Mauve alignment of JG068-like phage. A sequence alignment that compares the nucleotide sequence of a *Burkholderia* phages (vB_Bcc_4a) with JG068. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

The *Burkholderia* phage vB_Bcc_11b shares a blastn sequence homology (8.406 kbp) to *Burkholderia* phage KS5. Figure 5.9 shows the mauve alignment between the two phages where the sequence homology is notably more apparent and widespread across the genome. A region where the vB_Bcc_11b omits genes is seen in the KS5 genome at around the 26 and 29 kbp in the reference KS5 genome. The KS5 genome at 9 kbp shows a mobile genetic element flagged by repeat regions on either side of the genome which is not present in vB_Bcc_11b. The KS5 genome also has a putative exonuclease like gene between 23 – 25 kbp which is not present in vB_Bcc_11b. This inadvertently means that the vB_Bcc_11b phage is around 5 kbp shorter. The two regions where the vB_Bcc_11b has a putative gene not present in the KS5 genome is a hypothetical protein at 6.3 kbp and a tRNA anti-like protein at the 14.6 kbp region of the genome.

Figure 5.10 shows the mauve alignment of 3 phages (vB_Bcc_4b, vB_Bcc_16a and vB_Bcc_39a) that according to the blastn similarity search showed only one of the 3 having over 11 kbp region of sequence similarity to KS9. However, the alignment shows a red block of nucleotide sequence which shares the most homology between all the phages. vB_Bcc_4b and vB_Bcc_16a display the most sequence homology to each other. The conserved regions between all the phages are the tail head capsid terminase, tail tip fibres, endolysin and holin protein. Another region conserved in all the phages is the pink block and is identified in a helicase.

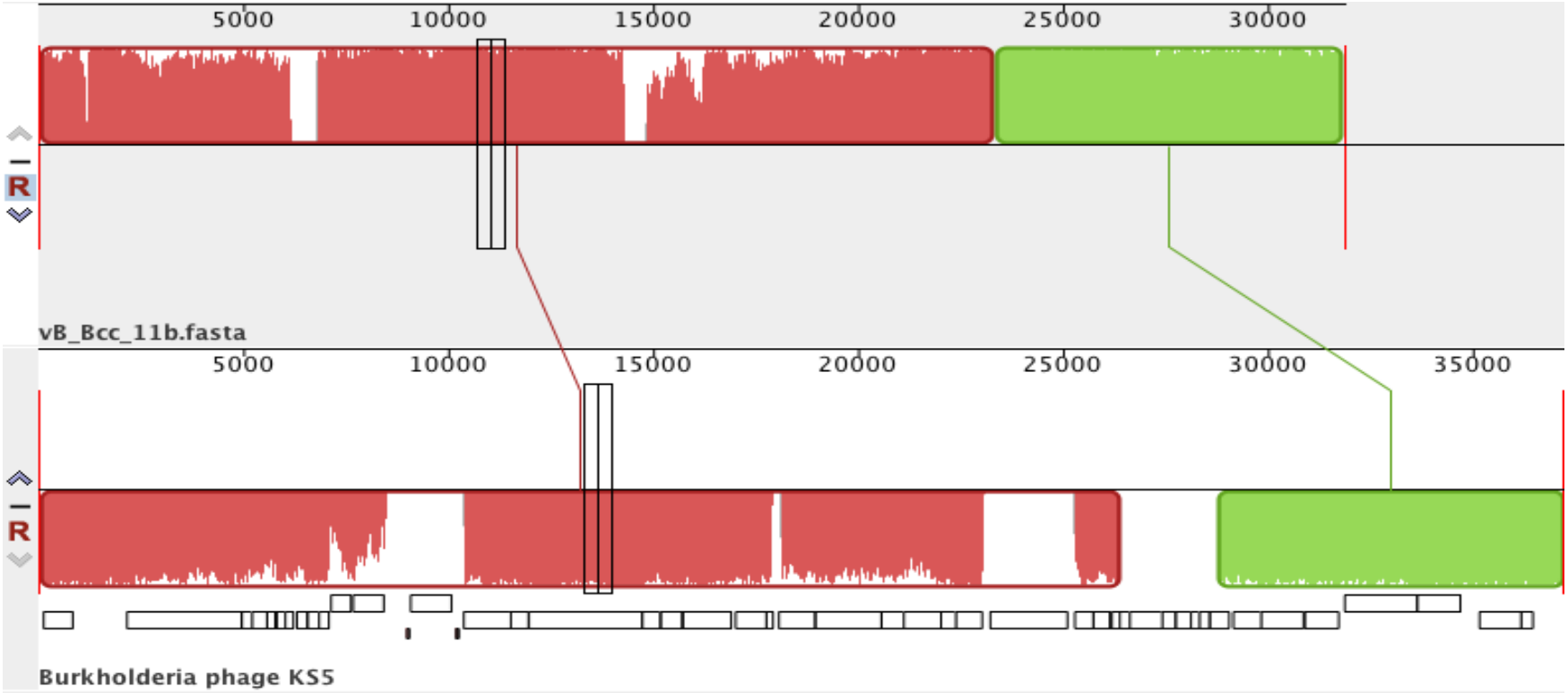


Figure 5.9 Mauve alignment of KS5-like phage. A sequence alignment that compares the nucleotide sequence of a *Burkholderia* phage (vB_Bcc_11b) with KS5. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

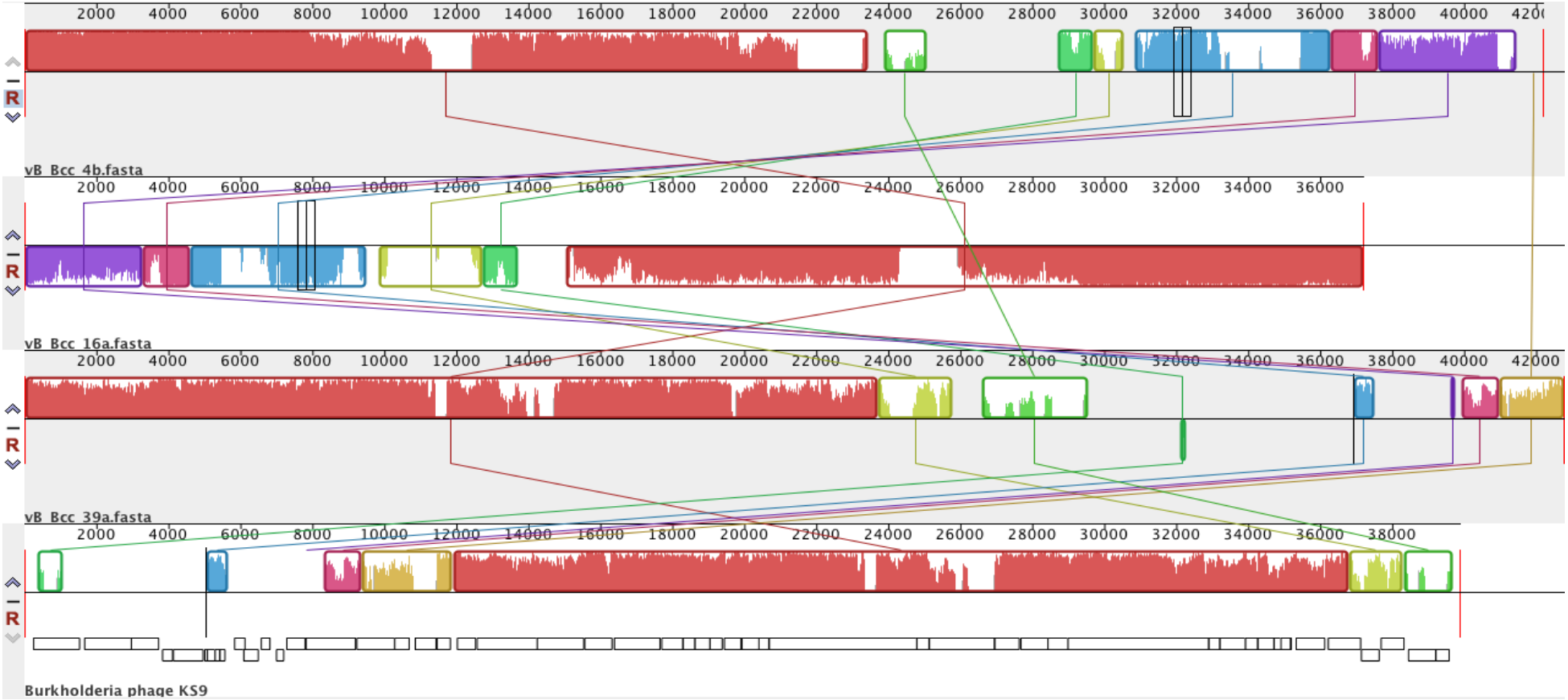


Figure 5.10: Mauve alignment of KS9-like phage. A sequence alignment that compares the nucleotide sequence of 3 KS9–like (vB_Bcc_4b, vB_Bcc_16a and vB_Bcc_39a) *Burkholderia* phages with KS9. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

Three *Burkholdria* phages shared very high sequence similarities to the phage KS10 (figure 5.11). There were a few bp variations predominantly at the right end of the genomes.

The phage vB_Bcc_31a displayed a 0.770 kbp match to Phi1026b. vB_Bcc_31a is a larger phage than Phi1026b (figure 5.12). Based on the mauve alignment the majority of this homology is on a single putative DNA methylase.

The vB_Bcc_15a phage shares sequence homology to PhiE12-2 which appears to be scattered across the whole genome (figure 5.13). The blastn sequence similarity showed a 2.606 Kbp homology (77%) between the phages. Sequence homology is identified between a small terminase, major capsid, capsid scaffolds, terminase ATPase, phage portal, addiction module antidote and killer protein. All the regions in the vB_Bcc_15a where no homology is seen to the reference PhiE12-2 are hypothetical proteins.

The phage vB_Bcc_43b shares a 0.751 kbp sequence homology to PhiE125 (figure 5.14). This similarity is seen in the centre of the genomes predominantly, the first gene is a putative DNA adenine methylase and the second region is similar to a plasmid region.

The *Burkholderia* phage vB_Bcc_32a shares a 1.048 kbp match to a *Ralstonia* phage Phi RSA1 (figure 5.15). This match is predominantly over a putative transposase gene at the 27 kbp region on the RSA1 phage. The genomes share homology to a putative terminase gene at 8 kbp region of the Phi RSA1 genome.

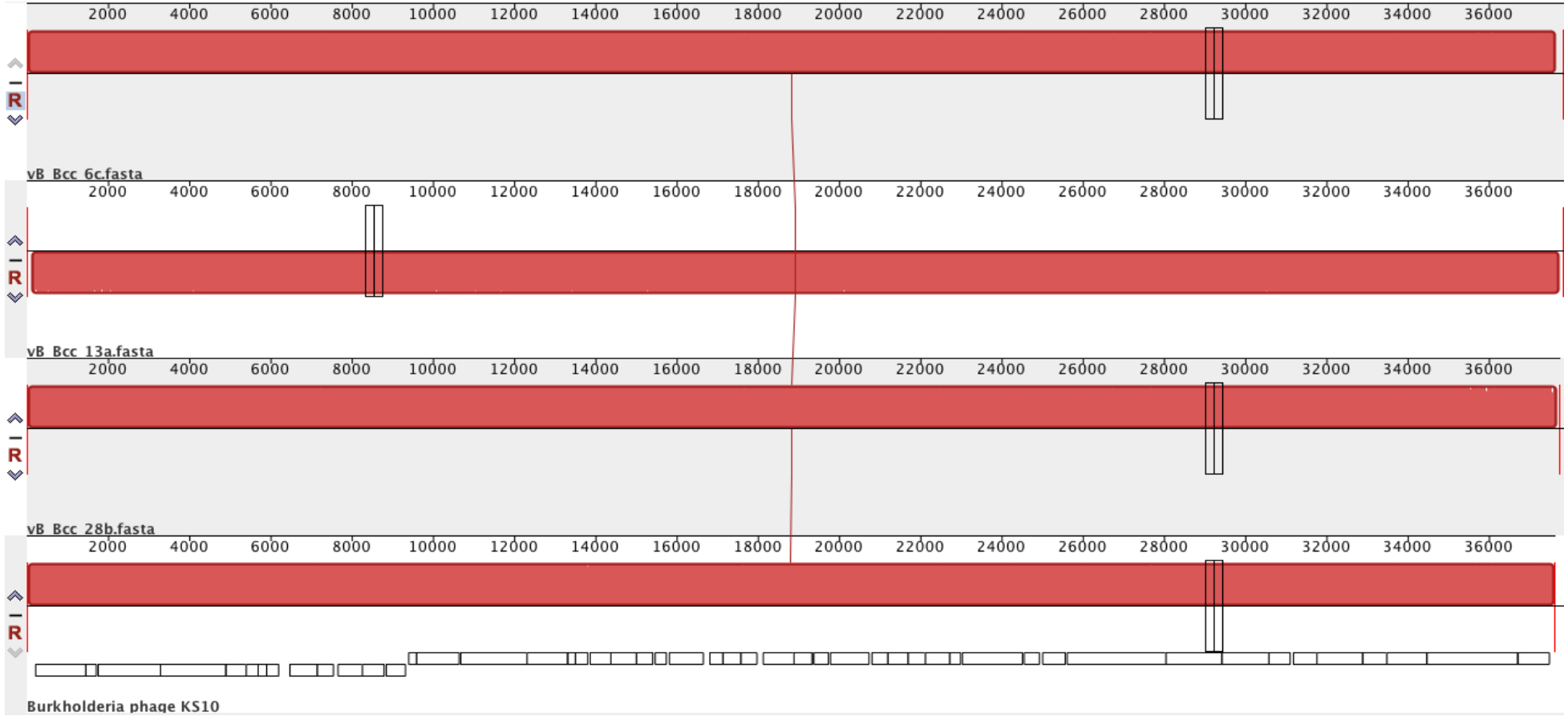


Figure 5.11 Mauve alignment of KS10-like phage. A sequence alignment that compares the nucleotide sequence of 3 KS10–like (vB_Bcc_6c, vB_Bcc_13a and vB_Bcc_28b) *Burkholderia* phages with KS10. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

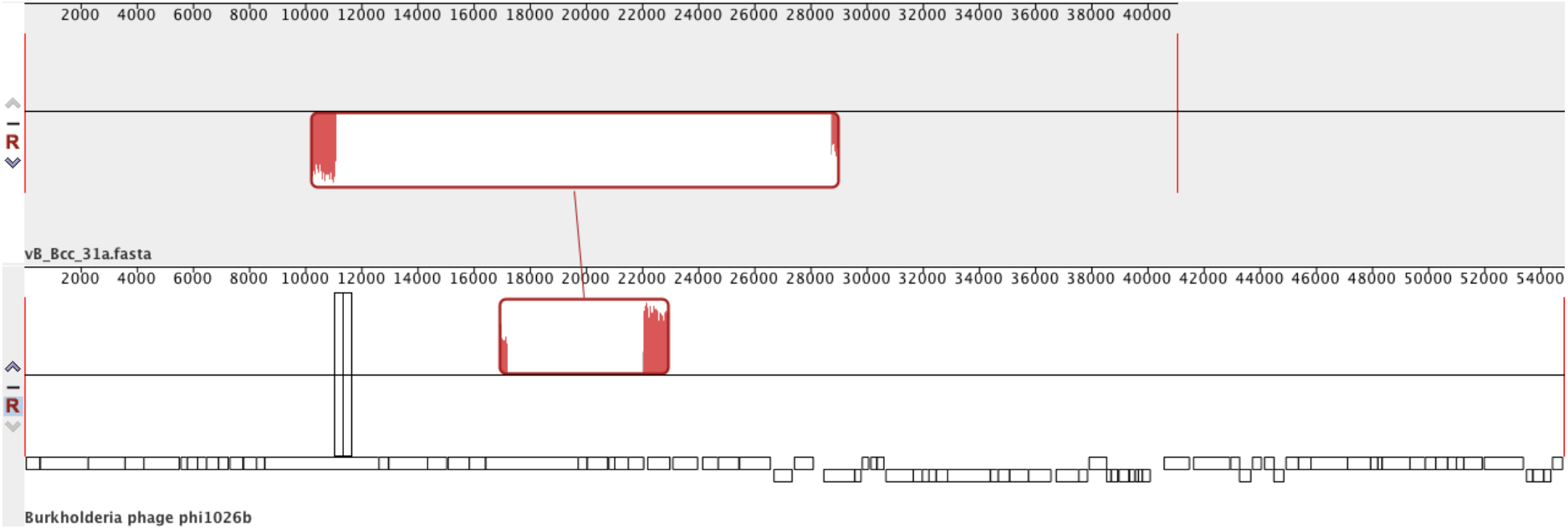


Figure 5.12 Mauve alignment of Phi1026b-like phage. A sequence alignment that compares the nucleotide sequence of a *Burkholderia* phage (vB_Bcc_31a) with phi1026b. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

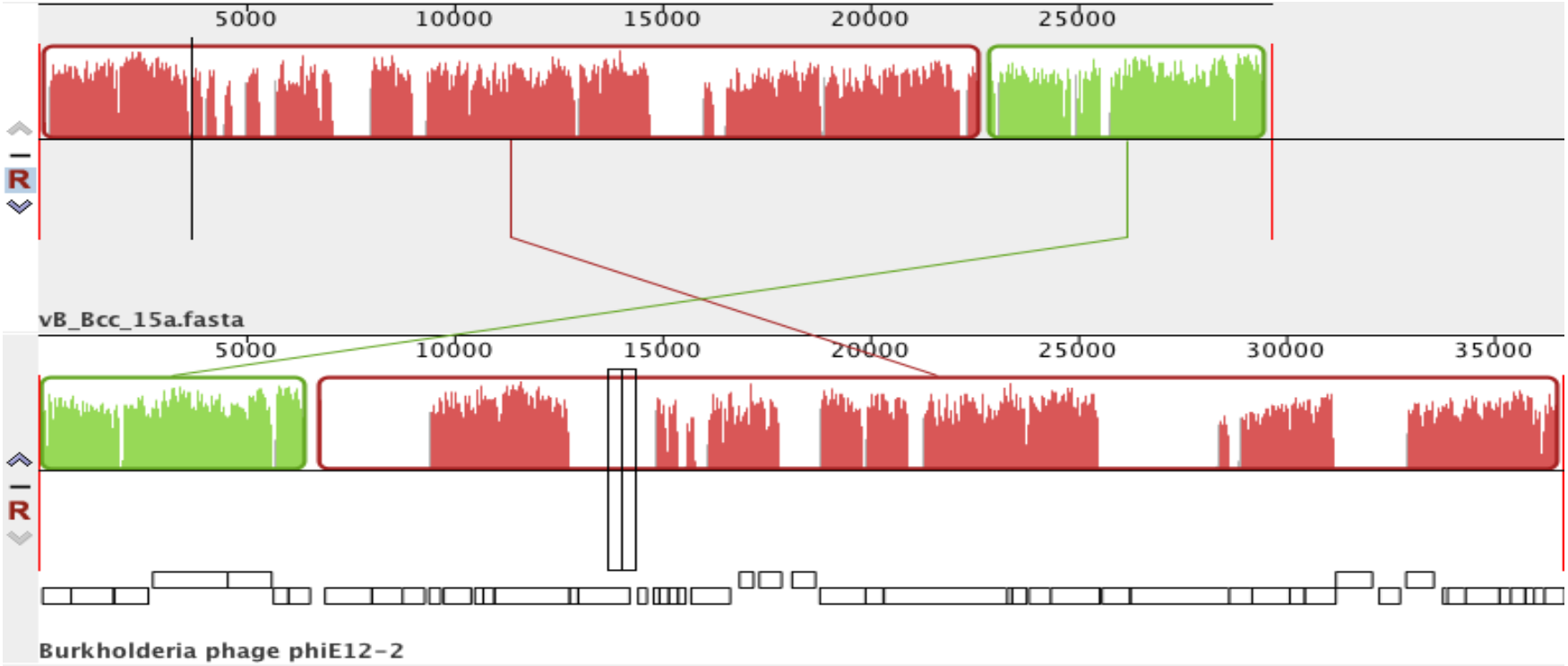


Figure 5.13 Mauve alignment of PhiE12-2-like phage. A sequence alignment that compares the nucleotide sequence of a *Burkholderia* phage (vB_Bcc_15a) with phiE12-2. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

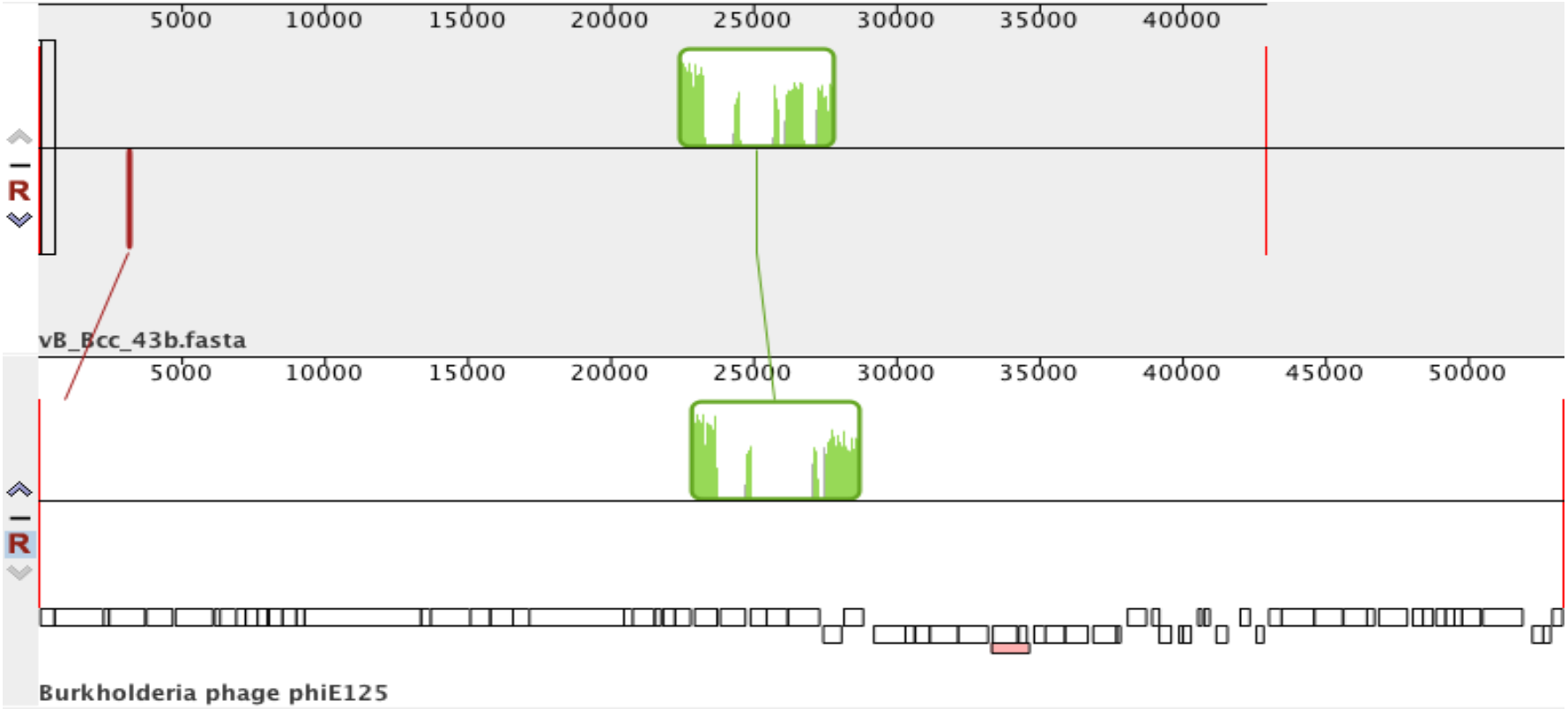


Figure 5.14 Mauve alignment of PhiE125-like phage. A sequence alignment that compares the nucleotide sequence of a *Burkholderia* phage (vB_Bcc_43b) with phiE125. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

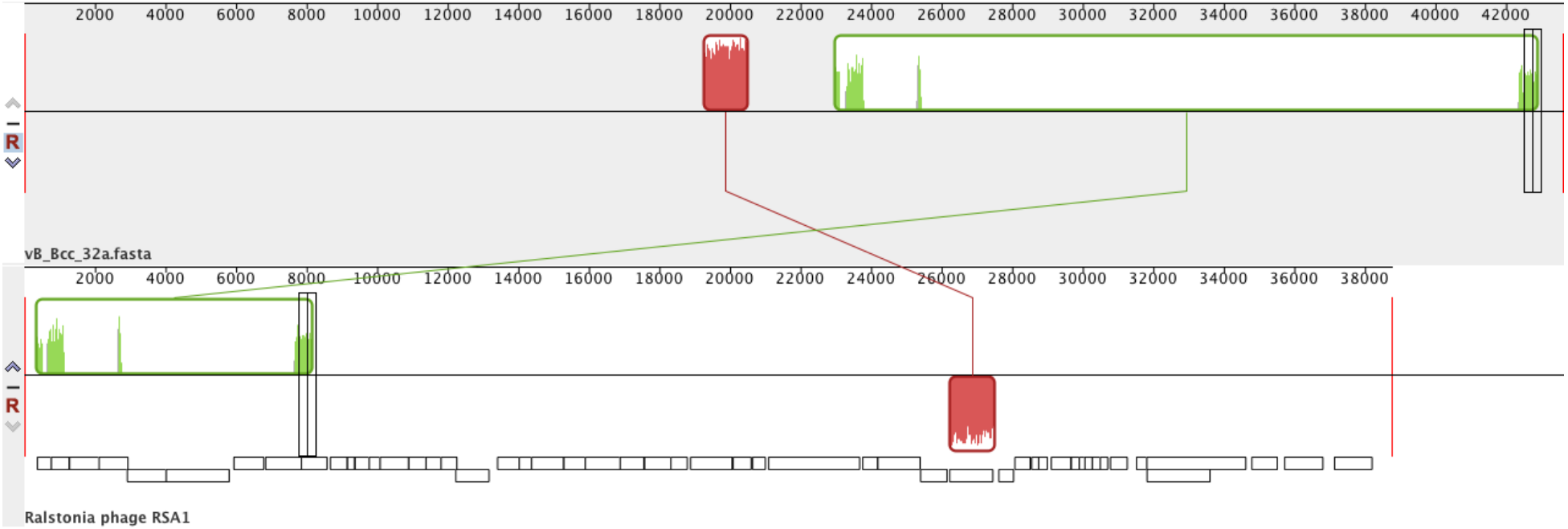


Figure 5.15 Mauve alignment of RSA1-like phage. A sequence alignment that compares the nucleotide sequence of a *Burkholderia* phage (vB_Bcc_32a) with phiRSA1. The coloured backbone makes it easier to see where there are shared sequences matches. The alignment was generated using Mauve v2.3.1 using progressive alignment.

5.3 Discussion

5.3.1 Cross infection of the temperate phage

The results presented this study detail induction, DNA extraction of genomic phage DNA, sequencing and characterisation of BCC phages. We identify that every strain of Bcc isolated from the CF lung environment was shown to harbour at least one inducible temperate phage when stressing the cell using the antibiotic norfloxacin. The cross infection study mirrors the approach and objectives of chapter 3, in that the approach looks to map the infection profile the potential mixed lysate populations of phage. Interestingly here we show that 68% of these phages were able to re-infect their originating Bcc host and therefore do not conform to the lambda temperate phage model of inhibition to superinfection. Superinfection in the context of bacteriophages is a re-infection of the same host bacterial cell by an isogenic phage (Adams, 1959).

From the data presented 15 carry only one phage, the other 5 are polylysogens. In this study we showed that the Bcc isolates displayed varying susceptibility to phage infections and the Bcc phages showed varying infection profiles. We showed similar findings with the Pa isolates and their phages. The limited clinical data associated with the Bcc samples unfortunately meant that this study could not extrapolate any other details from the cross infection data

However, it can be inferred that the variation in the phage/host dynamics seen in the cross infection study, could be associated with the temperate phage resistant strategies the bacterial host evolve, in order to evade phage lysogenic infections (discussed in chapter 1.6.6). This study showed that the average percentage infection between the phages from genomovar III A and multivorans species remained comparable at 64% and 65% respectively.

5.3.2 Bcc phage genome isolation and characterisation

Bcc phages were induced using the protocol described in chapter 2.3.1 and bacterial DNA and RNA was enzymatically cleaved and isolated using the method described in chapter 2.6.2. The khmer toolkit was used (as described in chapter 2.6.5) to remove any erroneous k-mers of k length < 32 and by means of extracting out k-mers based on abundance and count, cross over bacterial DNA was significantly depleted. The k-mer graphs pre and post sequence extraction for the Bcc phages can be found in table 5.1. What was immediately evident from the Bcc k-mer graphs compared to the Pa k-mer graphs fewer peaks were observed. These translated into fewer phages as per Bcc isolate. This had a direct effect on the assembly of these phages, where fewer issues with chimeric assemblies were seen and thus made it easier to validate complete phage assemblies. From the panel of 47 bacterial isolates 26 Bcc carried inducible phages. The phage genome assemblies were presented and their circular genome maps can be seen in appendix 4. These phages and their closest taxonomical sequence similarity are tabulated in table 5.2. The first limitation of this method is linked to the low sequence yield which could be associated to the phage titre in the lysate or number of raw reads per sample used for assembly. Samples that had partial assemblies were not shown in this study and thus only 26 Bcc phage genomes are presented.

Of the 26 phages presented, seven showed sequence homology to Bcep176, 4 to BcepMu, 3 to KS10 and KS9 and 1 each to Bcep781, DC1, JG068, KS5, Phi1026b, PhiE12-2, PhiE125, PhiRSA1 and J2315 ch2 like phages. Although only 26 complete Bcc phages were identified they appear to be more dissimilar as 13 different types were identified based on sequence similarity. All the phages are of the *Caudovirales* order and are members of the *Siphoviridae*, *Myoviridae* and *Podoviridae* family based on their sequence homology. Blastn searches offer little information about the phage genome architecture bar identifying sequence homology. Three KS10 like phages (figure 5.11) and the phage sharing homology to a *Burkholderia* J2315 chromosome two (figure 5.7) exhibits nearly 100% sequence identity matches. All the other phages in comparison

have varying sequence homology. Thus, the mauve alignments offer a greater depth of information concerning the genome architecture, identical sequence homology and percentage homology. Figure 5.9 displays a vB_Bcc_11b phage that shared according to the blastn search 8.4 kbp match to KS5. However, the mauve alignment revealed a greater degree of sequence conservation and homology seen between the two phages. Using the mauve alignment, it could be seen where the vB_Bcc_11b was absent from a mobile genetic element and exonucleases gene which were present in KS5 and instead a putative tRNA anti-like genes was present in vB_Bcc_11b genome at this position and absent in the KS5.

This study shows that the phages in Bcc do exhibit genomic and potentially morphological diversity. Unfortunately Bcc temperate phage are not well characterised and therefore little is known (Langley et al, 2003; Nzula et al, 2000). Bcc strains have been through taxonomical revisions over the last few decades and some phage pre-date this revision. Examples include phage CP1 (Cihlar et al, 1978) and CP75 (Matsumoto et al, 1986). Both of these authors refer to the bacterial strain as *Pseudomonas cepacia*. This confusion within the taxa of the bacteria may have repelled researcher away from studying these bacteria and their phage.

Bcc strains are used as bioremediation and as fungicidal in eastern countries and are seen as ‘good’ bacteria. However, our growing understanding of this strain being opportunistic and problematic in CF and chronic granulomatous disease in the western countries has given them a bad ‘vibe’ (Mahenthiralingam et al, 2002). A virulent Bcc ET12 strain has been shown to be multidrug resistant and to some of the strongest Bcc antibiotics, meropenem and ceftazidime (Nzula et al, 2002). The predominant research focuses on lytic phage based therapy as this type of research can offer direct impact on disease and clinical management. A study looking at 20 Bcc isolates identified that lysogeny is common in Bcc strains as they went on to identify 14 phages (Langley et al, 2003). This is a common trait in most dsDNA temperate phages. This method however only involved screening of spontaneous phages and this is where this study is also novel

in that it is using a chemical inducer in the form of an antibiotic NFLX. The research presented here could potentially increase the available pool of Bcc temperate phage genomes.

This study shows potentially eighteen novel phages that either share weak and or conserved sequence homology to phages. We show that four phages (figure 5.4) are slightly larger than the BcepMu like phage. This would conform to the hypothesis of this phage adding ~ 2.5 kbp of the host bacterial chromosomal DNA into its capsid due to its unique packing strategy (Summer et al, 2004). This also highlight that these phages still need to be evaluated for their genomes entirety.

Based on the aims of this chapter it can be concluded that Bcc phages exhibit varied infection profiles and can harbour more than one temperate phage in their genomes. However, in comparison to Pa sample the Bcc samples do not have a high diversity of phages in their lysate, making downstream sequence analysis comparatively easier to analyse. The study will add sequence data associated to Bcc temperate phages to online databases, that would offer improve downstream comparative genomic analysis.

6. Confirming the bacteriophage adherence to mucus (BAM) model on temperate bacteriophage and the impact of choosing the correct *de novo* assembly algorithm for metagenome assemblies.

6.1 Bacteriophage adherence to mucus (BAM) Model

In the lungs mucus acts as a barrier for protection against pathogenic infection of the epithelia. It achieves this by forming a layer between the lumen of the airways and the primary epithelial cells (Barr et al, 2013a; Hansson, 2012). Mucus is composed of mucin, hydrophobic glycoproteins (Cone, 2009; Rose et al, 1987) and other macromolecules that are also known to be present depending on which organ (Kim & Ho, 2010). The Bacteriophage Adherence to Mucus (BAM) model proposes that lytic phages adhere to carbohydrate residues present within the mucus layer, and provide a layer of immunity to incoming bacteria as the tail fibers adsorb to the bacterial surface, infecting incoming bacterial cells, multiplying and therefore lysing the cells. This therefore offers another layer of defense and immunity against certain bacteria trying to colonise the mucus layer. The cell debris of the lysed bacteria and excess phages are expelled with the mucus (Barr et al, 2013a).

The BAM model is facilitated by structurally displayed carbohydrate adherence domains, such as the immunoglobulin (Ig)–like domain present on the highly immunogenic outer capsid (HOC) of *Myoviridae* phage (Fraser et al, 2006). Structural proteins with associated Ig-like domains have been found in approximately 25 % of the sequenced dsDNA phages of the *Caudovirales* order, demonstrating their ubiquity and their potential importance in aiding phage survival (Fraser et al, 2006). Ig-like domains have been found on structural genes of the *Siphoviridae*, *Myoviridae* and *Podoviridae* bacteriophages (Fraser et al, 2006; Tariq et al, 2015). Fibronectin is a 440 kDa is an example of a glycoprotein that is complementary to an integrin. Integrin's are transmembrane receptors that adhere cell to cells (Pankov & Yamada, 2002). Fibronectins are involved in a wide range of cellular interaction including cell-cell adhesion, migration and cell differentiation (Greiling & Clark, 1997; Valenick et al, 2005). Bacterial Ig-Like domains are found on bacterial cell surfaces and have Ig like folds which

interact with bacterial cell adhesion molecules (intimins) (Kelly et al, 1999). Fibronectin type 3 domians have been seen on tail fibres of *Siphoviridae* and a type 2 bacterial Ig-like domain (Big_2 domains) on major head protein (MTP) of the same family (Fraser et al, 2006). Such structural domains have been seen to be conserved for phage propagation in laboratory settings as they mediate interactions between the phage and its host cell (Fraser et al, 2007; McMahon et al, 2005). Conservation of these domains increase infectivity, binding to degrade polysaccharides and the domains can be horizontally transferred (Fraser et al, 2007). Figure 6.1 shows a schematic diagram of this process.

In chapter 3 of this thesis the cross infection study offered a large data set showing adaptation and the change in dynamic interaction between phages and their hosts as chronic respiratory disease progresses. As there is a difference in interaction based on the phage infection profiles, then it is pertinent to look at other traits that correspond to the environment of the chronic lung and disease progression. As there is thick viscous mucus, it was pertinent to hypothesise that carrying a complex carbohydrate binding motif would offer a selective advantage for temperate phages too. This affinity, holding the phage at the mucus would also enable infection of incoming bacteria to the environment and may aid the addition of genetic variation to the community that benefits both the phage and the bacterium.

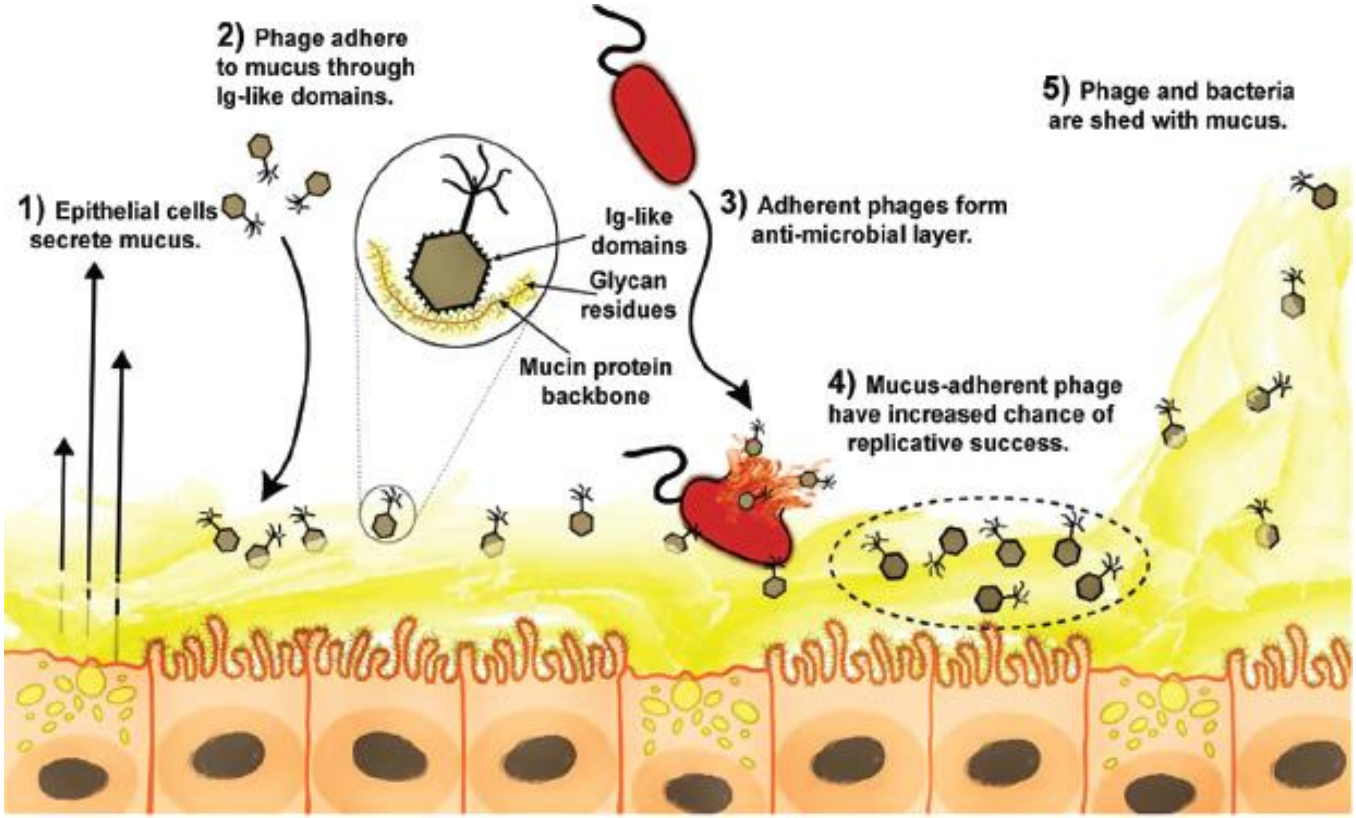


Figure 6.1 Bacteriophage adherence to mucin (BAM) model. The diagram show the lytic phage mediated immunity. Lytic bacteriophages adhere to the mucus secreted and utilise the environment to infect any incoming bacteria. The result of successful lysis increases the phage titre, excess phages are cleared out with the mucus maintain an equilibrium in the lung environment. The image was taken from Barr *et al* (Barr et al., 2013).

In chapter 3 it was demonstrated that every strain of Pa isolated from the lung environment had inducible prophage. The study also identified that phages chemically induced from a CF Pa background when spotted onto BR Pa strain backgrounds displayed higher infection rates when compared to phage induced from BR Pa background. Interestingly a higher number of CF Pa phages could re-infect their originating hosts compared to the BR phages. This finding goes against the global convention of temperate phage infection, as the prophage repressors function is to prevent superinfection (Kholodii & Mindlin, 1985). This suggests that the CF Pa phages and hosts have adapted and hypothetically are more evolved with an ability to re-infect their originating hosts in this environment. This offers variation through gene addition and the impact this has. Fogg et al 2011 illustrated that subsequent infection of isogenic phages with Stx-phage 24B (Fogg et al, 2011), promotes superinfection (Allison et al, 2003) but also increases toxin production through multiple copies of the gene present with polylysogeny (Fogg et al, 2012). This adaptation is possibly more pronounced due to the lung environment being constrained, in the lower lung bronchi with pockets of infection alongside poor movement of the mucus and the ability of Pa to form biofilms. In the genome annotation of the Pa phage (appendix 3) we see genes associated with immunoglobulin (Ig)-like domains, these domains may be involved in protein-ligand and protein-protein interactions.

This study investigates whether the temperate phages isolated from the mucus rich environments of CF and BR patients' lungs support the BAM model and its clinical relevance. We here propose a different strategy for the BAM model for phages to disseminate across their host range. These observations led to the hypothesis that these domains may aid in both the adsorption of phages to their bacterial hosts and the mucus layer under certain environmental conditions (Fraser et al, 2007). This study hypothesizes that temperate phages may use the mucus barrier as a way of infecting incoming bacteria which may drive gene exchange and add another level of adaptation and evolution. It also may be a way of increasing genetic heterogeneity in a population

of bacteria in late stage chronic lung infections that are traditionally thought to be somewhat clonal. This report compares the inducible temperate phages of Pa found in the lungs of patients with CF and BR using a metagenomic approach and studies the complexity of functionality the phage accrues through their continual adaptation and evolution in the chronic lower lung.

In order to characterise the Pa phage further we isolated their DNA and sequenced them on the Illumina MiSeq. In this study the mixed viral communities induced from each isolate were harvested, DNA extracted and sequenced were assembly using was attempted using three assembly algorithms; SPAdes v3.1.0, IDBA-UD v1.1.1 and VelvetOptimiser v2.2.5. Using these 3 assemblers allowed comparison and whether each offered difference in incidence or type of Ig binding motif present from differences in algorithm. In this study *de novo* assembled phage genomes were searched for putative carbohydrate adherence domains (CAD) (Tariq et al, 2015). This was tested to assess the relevance of the bacteriophage adherence to mucus (BAM) model initially proposed by Jeremy Barr (Barr et al, 2013b) in temperate phages. For the purpose of this study we used 45 BR samples and 47 CF samples.

6.1.1 Genome assemblers

There are a number of paid and open source *de novo* genome assemblers currently available for assembling contigs from Illumina based paired end reads. The three assemblers used here in this study utilise the de Bruijn graph assembly algorithm. This algorithm helps separate repeat regions as true repeats first proposed by Idury and Waterman (Idury & Waterman, 1995). The assemblers also utilises the error correction step added later by Pevzner that removed errors in the reads generated by the sequencer, a major problem for Sanger reads from the first generation sequencer at the time, as they had high error rates (Pevzner et al, 2001). PacBio presents this problem today with <Q10 accuracy (Quail et al, 2012). The three assemblers explored in a chapter

of this study are SPAdes, VelvetOptimiser (using Velvet) and IDBA-UD, all are open source.

The SPAdes assembler was first launched in 2012 by St Perterburg University. Since its launch it has been well maintained and is much easier to install and use compared to other assemblers. SPAdes is able to utilise reads from Illumina, Ion Torrent and can give hybrid assemblies from Oxford Nanopore, PacBio, and Sanger platforms. SPAdes was ranked well on the GAGE-B paper which looked at several popular assemblers and how they compared when assembling bacterial genomes (Magoc et al, 2013). A study compared SPAdes to eight other assemblers and found that it was the best at assembling *Caulobacter henricii* a GC-rich genome which can be problematic (Scott & Ely, 2015). Bacteriophages have also been shown to have GC-rich genomes, implying SPAdes as a good choice to assemble bacteriophage sequence data from Illumina MiSeq sequencer platform. SPAdes has four main steps that help assemble contigs from uneven coverage and to prevent chimeric assemblies. The steps involve firstly creating an assembly graph, biread adjustments, construction of pure assembly graph and finally contiguous sequence contruction using BWA. BayesHammer tool is used to improve error in contigs contruction.

Velvet also uses the de Bruijn graph to assemble *de novo* assemblies. VelvetOptimiser was developed by S. Gladman and T. Seemann and their group at Victorian Bioinformatics Consortium. VelvetOptimiser is a multithreaded perl script which helps to automate three parameters k, -exp_cov, and -cov_cutoff part of the velvet de novo assembler. Once you set your start and end k-mer lengths you choose the k-mer increments to be searched and VelvetOptimiser tries to optimise the parameters mentioned above. The k parameter optimises the best hash value given from the value ranges. The –exp_cov parameter helps assemble contigs truer to the size. If the –exp_cov is set to low the contigs sizes will be to low and if set to high you will force contigs together thus it is crucial to get the right coverage (Homolog.us, 2016). The longer the contigs the more likely the coverage will drop and the coverage cut-off

parameter allows the assembler to assemble longer contigs above the coverage cut-off set. Velvet uses two commands to assemble contigs velveth and velvetg. Velveth generates hashes for the reads and constructs roadmaps to the reads. Velvetg uses the information generated by velveth and build the de Bruijn graph, runs the error corrections and builds the contigs for each k-mer value set using the parameters in the VelvetOptimiser command.

IDBA-UD is a frequentative de novo assembler that uses de Bruijn graphs to assemble its contigs. This assembler is an extension of IDBA in that it is designed for assembling contigs from data of uneven sequence depth. IDBA-UD uses low to high threshold of coverage cut-off and small to large k values. IDBA-UD can build long contigs as it reduces gaps and generates fewer branches in the graph. As the name suggests it is good for data set with reads of differing sequence depth. This suggests that it would be an ideal candidate for induced temperate phage of differing titres and thus varying sequence depth.

The aim of this chapter was to compare three assemblers based on the number of Ig-like domains identified, the number of contigs and largest contig assembled. A comparison of Ig-like domains identified based on the disease aetiologies would be compared and it is hypothesised that Ig-like domains would increase along disease longevity.

6.2 Results

6.2.1 Assembly prior to IG-like domain search

In order to assess the BAM model, we first assembled our phage metagenomes and with preliminary trials showed that the phages assembled in different orientations with vastly different contig sizes depending on the assembler used. In this study the first analysis was therefore to assemble the phage metagenomes using the different assemblers. The assemblers used were SPAdes v 3.1.0 (Bankevich et al, 2012), VelvetOptimiser v 2.2.5 and IDBA-UD v 1.1.1 (Peng et al, 2012).

The assemblies were compared using the quality assessment tool (Quast) (Gurevich et al, 2013). Table 6.1 tabulate the Quast analysis of each assembler and the assemblies of the phage metagenomes. This table shows the number of contigs assembled for each metagenome, the average size of those contigs and the largest assembled contig size. This allows inference on the depth of the data present. What is apparent is that the three-way assembler comparison revealed that SPAdes assembled larger contigs which correlated to a fewer number of contigs. This necessarily did not mean that SPAdes was the preeminent assembler. One method that this study utilises in order to justify choosing an assembler is to search for protein domains. Figure 6.1 and 6.2 compare the genome assemblies of the assemblers and compare them back to the disease states. Figure 6.2 illustrates that despite the assemblers varying in the frequency of identifying Ig-like domains the trend, to a certain degree remains the same, which is that as the disease state progresses with time the number of Ig-like domains increase.

	SPAdes 3.1.1			VelvetOptimiser 2.2.5			IDBA-UD 1.1.1		
Pa Sample No	N50	Number of contigs >=500bp	Largest contig (bp)	N50	Number of contigs >=500bp	largest contig (bp)	N50	Number of contigs >=500bp	largest contig (bp)
1	8330	14	25454	4554	15	11351	8274	15	11253
2	2064	43	3807	792	36	1718	2308	28	3802
3	8039	11	27661	3658	18	9286	2942	20	7053
4	5714	30	16984	760	26	1239	5713	27	16937
5	52800	2	52800	28349	6	37510	37577	3	40523
6	61649	3	61649	57714	10	57714	57517	4	57417
7	14035	14	37361	11114	18	37204	37280	18	37280
8	1909	49	5871	3569	17	11176	11439	17	11439
9	8363	15	24014	3560	12	9299	8274	14	11803
10	9016	14	20434	1677	28	5159	7726	6	20386
11	1606	8	2802	1519	8	2515	4822	3	4822
12	3603	29	7027	3521	29	7017	3646	22	6696
13	8330	17	20288	4878	14	15649	8307	13	20467
14	17035	41	38360	13243	58	38356	17717	46	38360
15	62077	6	62077	10313	17	20962	36064	8	37632

16	1670	46	7099	3525	18	9331	5161	17	11467
17	52664	2	52664	28259	3	37474	52623	2	52623
18	1366	63	4913	2021	22	4868	4698	16	8307
19	23619	4	29993		No contigs >500 bp		20028	6	28817
20	61562	6	61562	25865	16	37420	34762	7	37677
21	3055	20	11703	9300	14	11150	8274	15	9366
22	61772	7	61772	1364	50	2323	21201	5	26599
23	4681	21	8473	4666	21	8463	4765	11	8442
24	4019	6	4019	794	8	1864	3822	1	3822
25	10598	28	44244	10341	38	24798	9896	27	37577
26	37617	17	41525	10208	21	37479	24044	12	37559
27	39705	1	39705	39558	1	39558	39579	1	39579
28	3772	24	7654		No contigs >500 bp		8274	13	19484
29	10321	50	28830	12607	54	27562	13226	53	27631
30	11205	58	33124	5673	85	27617	7741	64	37496
31	26208	37	80678	15245	49	67147	18705	31	70475

32	17035	43	32899	8268	60	31650	17558	48	30584
33	39659	1	39659	19927	2	19927	25118	2	25118
34	61597	7	61597	1749	38	4913	23998	7	26599
35	17035	38	38239	11482	54	38244	17558	50	38047
36	37242	8	37654	28201	12	37474	25055	7	37576
37	8330	12	14034	3679	15	11087	8274	14	11163
38	3384	32	8039	1113	49	3288	4468	18	8007
39	2936	29	7842	3073	21	11263	5781	16	19353
40	2214	42	7645	3523	16	8460	5815	15	12799
41	12686	34	28104	6299	44	27593	9242	45	28071
42	28149	32	41380	27649	44	40129	26266	46	38243
43	3853	21	6849	3985	19	9873	2942	18	7317
44	52658	2	52658	31041	3	37504	37577	4	37832
45	32297	16	38297	38284	17	38284	38286	4	38386
46	10760	17	36835	9991	17	36766	10250	7	28792
47	15903	29	38353	15686	40	38341	15698	26	38341
48	3385	16	5970		No contigs >500 bp		39556	1	39556

49	4467	5	4467	2118	3	2118	4467	2	4467
50	17035	30	48272	8869	41	28497	16764	24	37577
51	26325	26	53682	14917	44	38255	15721	29	38116
52	8363	13	14150	3566	14	9457	8274	14	11198
53	50013	1	50013	41926	2	41926	49807	1	49807
54	21563	37	41593	19243	49	40329	17868	46	40398
55	14052	14	37359	4881	20	37204	8786	16	37021
56	17035	49	38328	13241	60	38328	17868	49	38316
57	4085	2	4085	812	1	812	3059	1	3059
58	3032	8	3249	900	9	3375	3032	4	3249
59	61772	1	61772	30843	4	30843	61582	1	61582
60	9128	6	9128	2595	8	4556	10076	1	10076
61	8089	14	25388	3729	17	11209	11167	17	11167
62	40441	2	40441	40414	7	40414	40352	2	40352
63	37715	1	37715	19523	5	19523	20272	2	20272
64	61772	1	61772	12408	6	17991	61609	1	61609
65	1193	57	5705	3760	16	9306	4071	17	11107
66									

67	886	6	3110	1449	2	1449	2529	1	2529
68	8089	17	24425	3563	14	11151	8274	15	11180
69	31654	26	90429	9645	38	31555	31625	26	37639
70	3712	7	4114	806	9	1065	5574	2	5574
71	594	1	594		No contigs >500 bp				
72	25633	2	25633	32516	2	32516	37350	1	37350
73	2257	41	6582	3702	16	15980	5815	15	19653
74	8237	11	14045	4194	19	9496	8307	14	11163
75	37660	11	52654	20956	21	31826	37510	9	52534
76	19009	21	35341	18985	13	31565	18969	10	35340
77	61772	1	61772	25201	3	26529	61570	1	61570
78	1806	45	6512	4009	16	10929	5784	18	9892
79	61772	1	61772	6092	10	19856	61609	1	61609
80	8122	13	25526	3497	14	8943	8307	14	10037
81	649	2	649		No contigs >500 bp				
82	2757	6	4408	2671	8	2713	4798	2	4798

83	2747	27	5652	1077	25	2160	3247	12	5624
84	28144	4	28144	23219	7	23219	28049	4	28049
85	12853	43	34949	10508	56	33590	14670	55	33749
86	8363	10	27106	3979	16	11175	8307	14	11122
87	1608	17	2327	1523	17	2287	1690	12	2287
88	26228	32	38158	13160	40	26010	17717	42	37876
89	8363	13	22288	1877	25	3681	4905	19	9710
90									
91	61772	1	61772	25193	3	26515	29967	3	30772
92	61772	1	61772	31735	3	31735	44032	2	44032
93	10045	10	15044	5913	11	11529	7128	8	23183
94	882	3	1348	575	2	575	1323	1	1323
Average	**20119.5**	**18.5**	**29063.5**	**11636.6**	**21.4**	**20462.0**	**18374.2**	**15.3**	**25739.4**

Table 6.1: Table of three-way assembly comparison between IDBA-UD, SPAdes and VelvetOptimiser. This table shows the assembly of each *Pseudomonas aeruginosa* (Pa) phage lysate genome sample. The Pa sample number refers to the genome sequence number assign to the Pa isolates. Three assemblers' comparison is shown; SPAdes v3.1.1, VelvetOptimiser 2.2.5 and IDBA-UD 1.1. The table shows the N50 score of the

assemblies, the number of contigs assembled larger than 500 bp and the largest contig (bp) present in the assembly. The table also takes an average of the assemblies across the samples omitting the samples with no value.

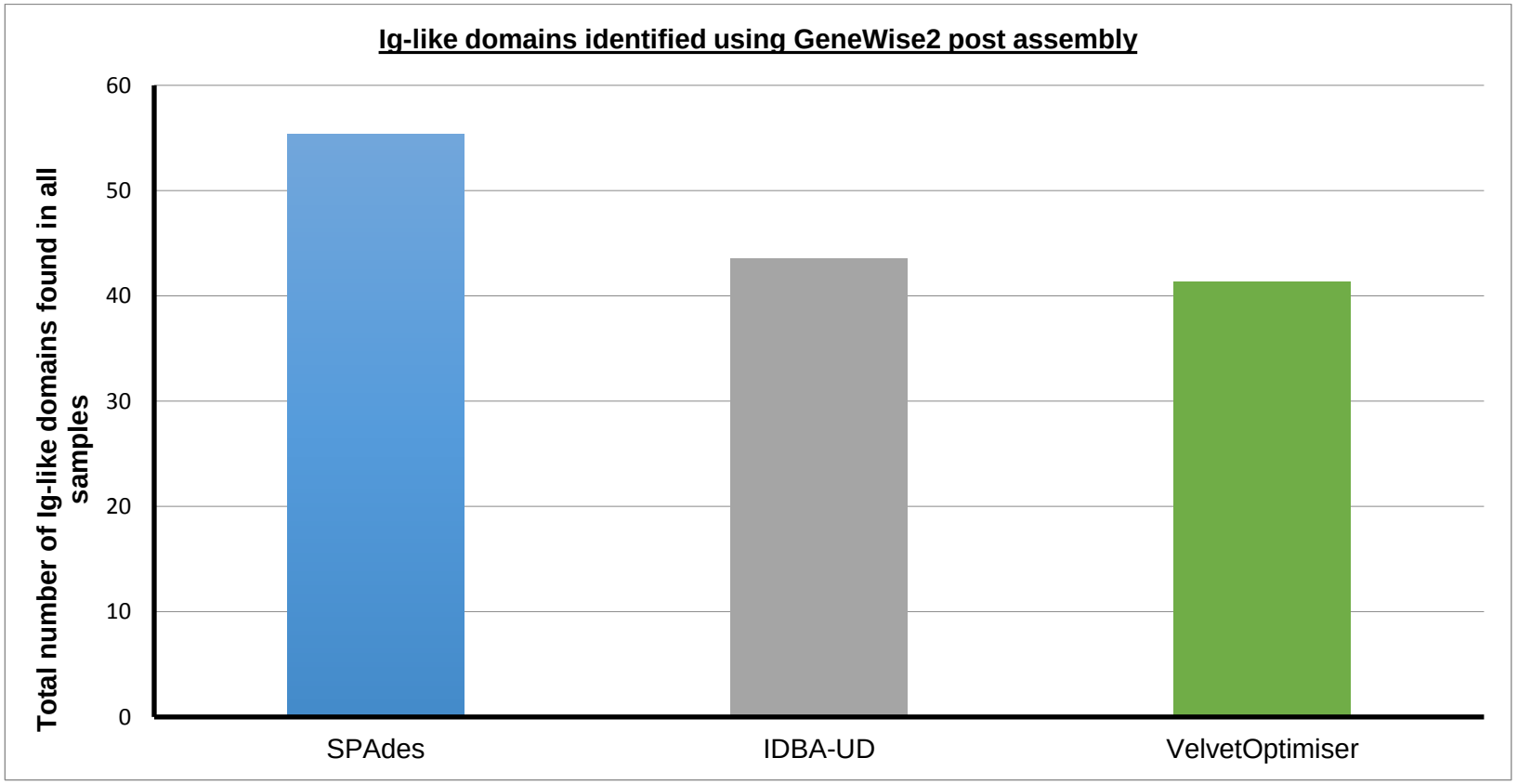


Figure 6.1: The total number of Ig-like domains identified as per assembler. Illustrates the total number of Ig-like domains identified using GeneWise2 post SPAdes, IDBA-UD and VelvetOptimiser Assemblies.

6.2.2 Identification of Ig-like domains

GeneWise2 was used to identify similarities between the assembled phages and the Pfams in table 6.2 (method chapter 2.10.1). Using the assembled contigs from the three assemblers this study identified the frequency of total Ig-like binding domains (Figure 6.1) and we stratified this further based on the 4 clinical disease states as previously mentioned (Figure 6.2).

To reiterate, the samples were separated based on following criteria < 10 years clinical diagnosis BR (17), > 10 years clinical diagnosis BR (28), Paediatric CF (10) and Adult CF (37). Big_2 domains were identified in 28 out of the 92 samples; Figure 6.3 shows this as a percentage (< 10 yrs clinical diagnosis BR 6%, > 10 yrs clinical diagnosis 32%, CF Paediatric 30% and CF adult 40%). Figure 6.4a depicts the domain architecture of the double Big_2 and He_Pig domains. Figure 6.4b illustrates the alignments of the Ig-like domains identified. The Big_2 domain was seen in 4 variations all having a higher bit score than the recommended cut-off of 25, which is considered significant. Using blastn on the open reading frame (ORF), the sequence alignment of the Big_2 domain was identified in a putative structural major tail gene.

We found that 42 of the phage samples had a He_Pig like domain all flanking a minor tail protein but in different genes. Figure 6.4b shows the He_pig alignment seen in over 70% of the 42 phage samples all other alignments of the He_Pig were seen to have a very low bit score and thus deemed to be insignificant.

All of the Big_2 domains were identified using the 6:23 algorithm and also the 21:93 algorithms. However, we found that only using the 21:93 GenWise2 algorithm a Big_3 domain was identified but was seen overlapping the first Big_2 domain by 89 amino acid, this could be linked to a frame shift (Fraser et al, 2006). This Big_2 and Big_3 overlap was seen in most of the samples.

An Fn3 domain was seen in 10 samples and was found on a putative phage tail assembly protein. Glyco_hydro_2 domain was identified but it was not seen linked to a potential structural protein.

SCOP Superfamily	PFAM Name	Accession Number
Ig	V-set	PF07686
	I-set	PF07679
	C2-set	PF05790
	C1-set	PF07654
	Ig	PF00047
	Ig_2	PF13895
	ICAM_N	PF03921
E-SET	Alpha_amylase_N	PF02903
	arrestin_N	PF00339
	arrestin_C	PF02752
	CelD_N	PF02927
	peptidaseC25	PF03785
	TIG	PF01833
	RHD	PF00554
	DUF291	PF03442
	Filamin	PF00630
	He_Pig	PF05345
Fibronectin type 3	FN3	PF00041
	tissue_fac	PF01108
	lep_receptor_Ig	PF06328
PKD	PKD	PF00801
	PPC	PF04151
	HYR	PF02494

β-Galactosidase/β - Glucuronidase	Glycol_hydro_2	PF00703
Cu, Zn Superoxide dismutase-like	Sod_Cu	PF00080
PapD-like	Pili_assembly_C	PF02753
	pili_assembly_N	PF00345
Invasin/intimin cell-adhesion fragments	Big_1	PF02369
	Big_2	PF02368
	Big_3	PF07523
	Big_4	PF07532
Clathirin adaptor appendage domain	Alpha_adaptin_C2	PF02883
Transglutaminase N-terminal domain	Transglut_N	PF00868
Cadherin-Like	Cadherin domain	PF00028
Actinoxanthin-like	Neocarzinostatin family	PF00960
CBD9-like	Domain of unknown function	PF06452
laminA/C globular tail domain	Intermediate filament tail domain	PF00932
Other Ig-like	C type Lectin	PF00059
	BACON	PF13004
	MucBP	PF06458

Table 6.2: Table of selected Ig-like domains. This table shows 40 Pfam databases used in GeneWise2 and adapted from Fraser *et al* 2006.

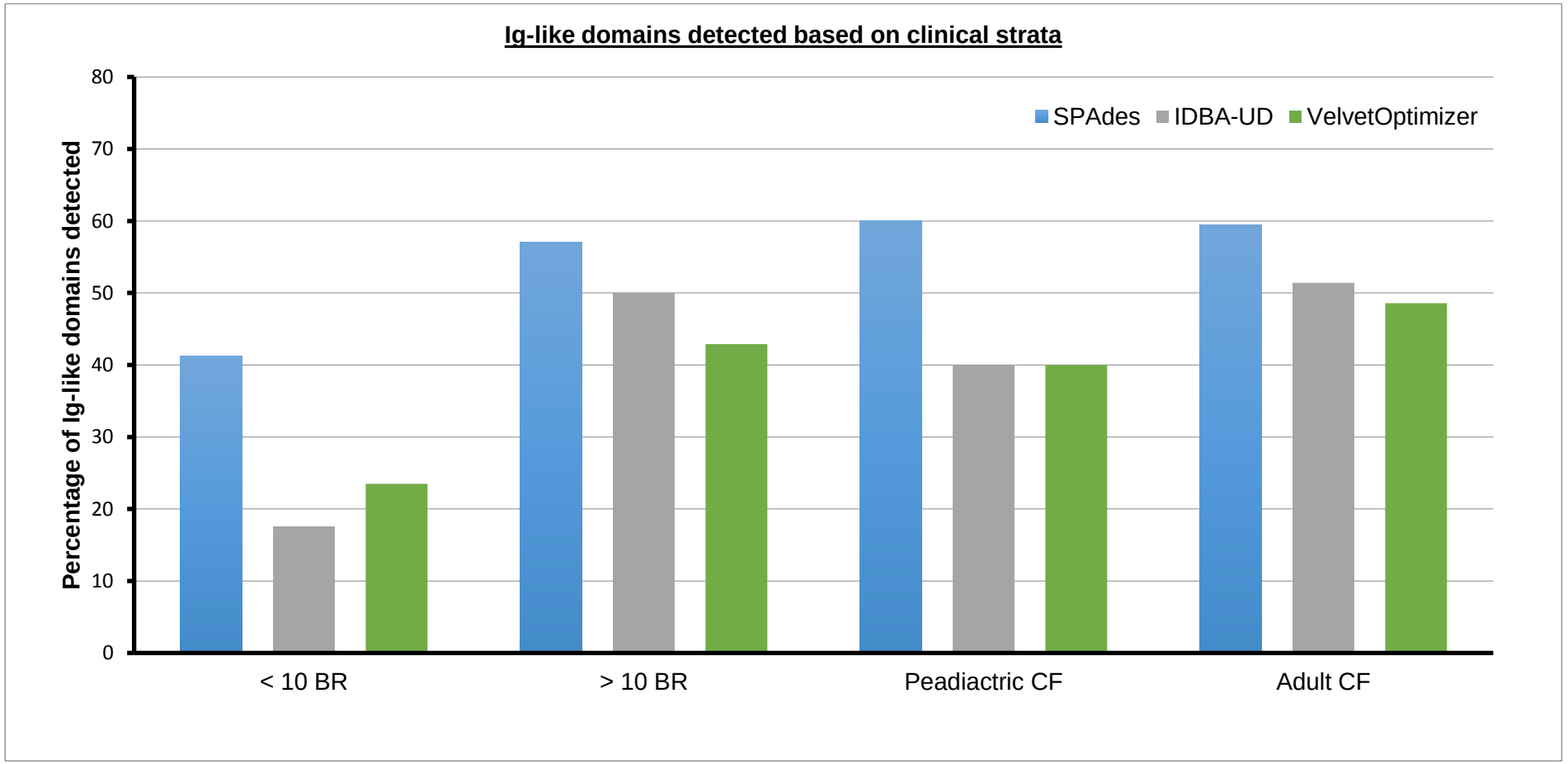


Figure 6.2: Clinical stratification of Ig-like domains identified as per assembler. Stratifies the total number of Ig-like domains identified by the three separate assemblies into the 4 clinical disease states.

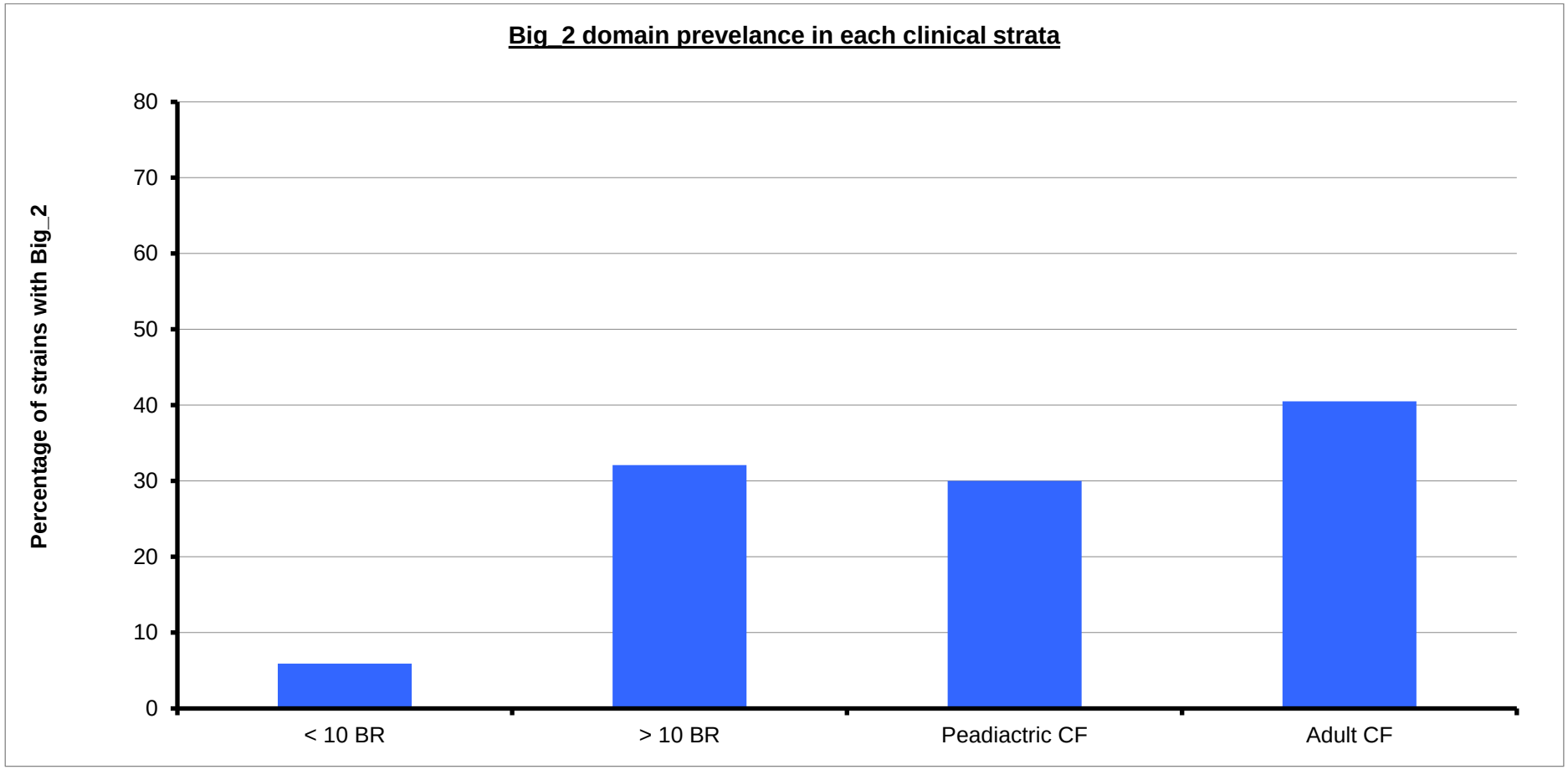


Figure 6.3: The total prevalence of Big_2 domain identified stratified against disease aetiologies. Concentrates on the percentage of Big_2 domains identified, as it attained the highest alignment score using GeneWise2.

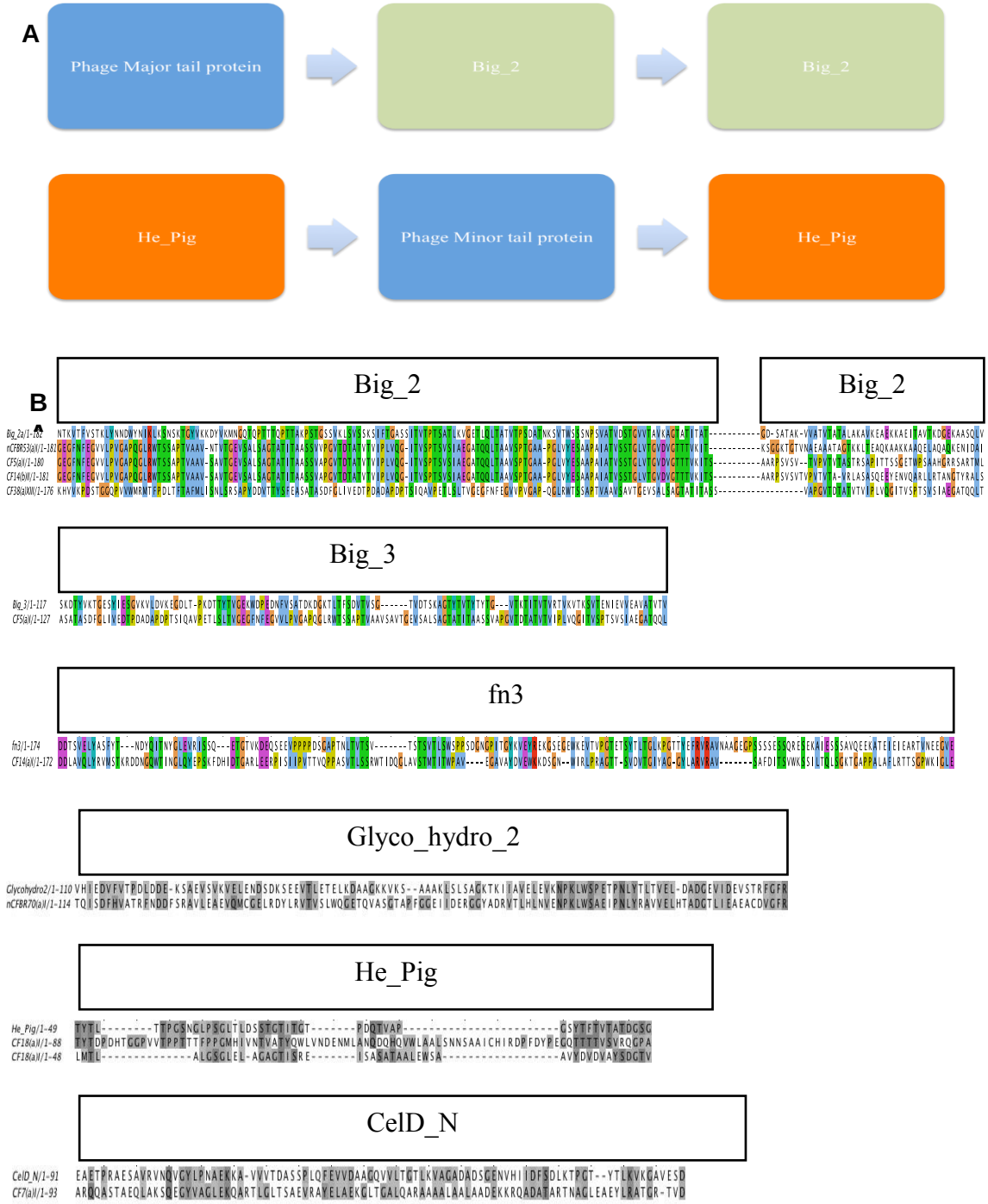


Figure 6.4: Domain architecture of Big_2, He_Pig and amino acid alignments.

Panel **A** shows the domain architecture of the double Big_2 domain that was identified in a subset of the phage lysates from the clinical samples, the second image is the proposed domain architecture for He_Pig. The He_Pig domains however, span over different genes adjacent to the Phage minor tail protein. Panel **B** shows the sequence

alignments of the Ig-like domains; Big_2, Big_3**,** Fn3, Glyco_hydro_2, He_Pig and CelD_N found in phage proteins. The sequence alignment of Big_2 is represented along with 4 variations that were seen in this study; all the samples have a high similarity to the Big_2 domain represented on the top line. Big_3 and Fn3 both have a lower sequence identity to the known sequence but this alignment was conserved between all the phages that were identified to contain these domains. The Glyco_hydro_2 domain was seen in only one phage sample (> 10 clinical diagnosis BR isolate). The He_Pig double domain was seen in 42 of the phage samples, 70 % of which have the amino acid alignment shown above.

6.3 Discussion

The lung of CF and BR patients is rich in mucus and has been referred to as mucoviscidosis disease when describing the gastro and respiratory tracts (Farber et al, 1943). The mucus acts as the first line of defense in the lungs and is secreted by epithelial cells (Rose et al, 2001). Mucus is mainly comprised of mucin glycoproteins (Rose et al, 1987). The Pa bacterial isolates isolated from CF and BR patients display varying mucoid phenotypes (table 2.2.1 and 2.2.2). Alongside the genome annotations of a few Pa bacteriophage showing putative Ig-like genes made it pertinent to compare the BAM model. The BAM model could suggest a means by which temperate phages infect across their host range in the lung. The main focus of the study was to identify specific carbohydrate binding domains including bacteriodetes associated carbohydrate often N–terminal (BACON) (Mello et al, 2010) as these were presented in the BAM model (Dutilh et al, 2014).

To investigate the BAM model 40 Pfam families of carbohydrate were compiled and assessed. The Pa phage metagenomes presented in this study no BACON domains in the panel of 92 mixed phage samples were identified. Conversely a Big_2 domain was identified in at least one of the phage metagenomes from the clinical backgrounds (3 pediatric CF isolates, 15 adult CF isolates, 1 < 10 year clinical diagnosis BR isolates and 9 > 10 year clinical diagnosis BR isolates). Equally a He_Pig like motif was identified at similar rates in both etiologies (CF and BR) however; it was observed that the incidence of He_Pig increased as the disease progressed over time.

In adult CF phage metagenomes 17 He_Pig like domains and in > 10 year Pa colonisation BR phage 14 He_Pig like domain were identified. This may suggest a role in phage adaptation to the environment whereby acquiring the ability to adhere to mucus aids the phage's chances to infect its host. He_Pig like domains were also identified in pediatric CF phage (4) and < 10 year clinical diagnosis BR phage (7).

This study also compared three open source genome assembly packages to assemble the 2 x 250bp paired end reads of each phage metagenome. We report that SPAdes and IBDA-UD are fairly comparable in performance (Figure 6.2 and Figure 6.3), while being able to identify a greater number of Big_2 domains using SPAdes assembled genomes. Statistically VelvetOptimiser produced the weakest assemblies based on the Quast analysis (Table 6.1) and the equivalent reciprocated when explicitly probing for Ig-like domains with GeneWise2. Nonetheless a VelvetOptimiser derived assembly contained an Ig-domain (PF00047) that both SPAdes and IBDA-UD failed to construct the associated contiguous seqeunce within their assemblies. Thus this study can stress a prerequisite of using multiple open source assemblers when trying to identify protein domains in metagenomes.

Protein similar to Ig-like domains have been identified on circa 25 % of all sequenced *Caudovirales* genomes, Big_2, I-Set and fn3 are three distinctive families which are only identified in *Caudovirales* phage (Fraser et al, 2006). Here we showed that using SPAdes over 60 % of our samples contained one or more Ig-like domain. This study confers to this as Big_2 and fn3 like domains were identified in a large subset of the Pa phages. In *Siphoviridae* and *Podoviridae* class phage these domains are often found to be associated with the major head protein (MHP), major tail and tail fiber proteins whilst in *Myoviridae* class phages they are situated on the highly immunogenic outer capsid (HOC), fibritin and baseplate proteins (Fraser et al, 2007).

When the BAM model was used to assess the lytic T4 phage it was observed to be associating with mucus via the head HOC protein (Barr et al, 2013a). Ig–like domains have been identified on the tail tube protein of *E. coli* bacteriophage λ, the precise purpose for these Ig-like domains on λ is not known but when these domains are trimmed, the phage has been seen to become more sensitive to temperature (Katsura, 1981; Pell et al, 2010). The Ig-like domains may have a more accessary roles due to their universal presence but they could be driving phage evolution to infect based on specific

environment being that they become temperature stable or increase their presence in mucus (Barr et al, 2013a; Pell et al, 2010).

In this study the Pa phages have a dual Big_2 domain downstream to a putative major tail protein (figure 6.4a). He_Pig domains were also seen in a paired motif which bordered a putative minor phage tail protein (figure 6.4a). He_Pig is an Ig-like domain found in hemagglutinin and cell surface proteins, making its involvement in the BAM model plausible. The core sequence variations were seen in the second Big_2 domain, figure 6.4a and b. All these domains had a bit score above the proposed minimum cutoff score of 25 and had a gap ranging from 18 bp to 6878 bp between the pair. This study and multiple domains posts the question does multiple motifs increase the selective advantage for the phage?

The BAM model postulates that the T4 phage displays Ig-like domains on the capsids and proposes a means for its adherence to mucus. In this study the putative Ig-like domains were seen to be linked with tail protein structures, as opposed to capsid proteins. This study hypothesis that the Ig-like domains for temperate phage with respect to the BAM model act as a mode of infection and phage dissemination rather than phage mediated immunity. The evolutionary pressure in the lung may be making the overall diversity of phage Ig-like domains very suggested by the increase in frequency of Ig-like domains identified as the disease state progresses in both CF and BR. Variation in Ig-like domains could evade detection of these domains with GeneWise2 (Fraser et al, 2006). However, making the use of multiple assemblers could help give a true representation of the incidence of protein domains. Based on the severity of the disease however, shows no correlation to the number of Ig-like domains found (clinical data, table 2.2.1 and 2.2.2).

This study can therefore infer that Pa phage induced from CF and BR patients can adhere to the mucus through interaction involving the Ig-like domains identified in the tail proteins. Arguably this is the largest study identifying Ig–like domains observed in the Pa clinical isolates. This study illustrates another possible evolutionary approach by

phage specifically to adapt in the lower lung. The Ig-like domains could help temperate phage infect neighboring host during spontaneous induction within the lung or aid attachment and infection of other cells in their direct environment outside of the lung.

The chapter shows that SPAdes assembler assembles temperate phages with the best N50 scores and helps identify Ig-like domains. The study also identified that Ig-like domains increased along disease longevity.

7: General Discussion

7.1 Discussion

The lower lung is a complex microbiological environment harbouring a polymicrobial community of bacteria, viruses and fungi (Purcell et al, 2014a; Sibley et al, 2006). However, rounds of inflammation, scarring of the lung, microbial colonisation and why interactions between the poly-microbial communities driving evolution resulting in a predominant species chronically colonising the lung remains unknown. What is known is that failure to eradicate bacterial infection in the lower lung reduces lung function and leads to respiratory failure. The chronic lung of CF and BR patient provides a niche where the pressures of this environment drive adaptation and diversification of these bacterial pathogens (Caballero et al, 2015; Davies et al, 2016; Winstanley et al, 2016). Pa and Bcc both readily colonise the chronic lung and are often the key chronic pathogens due to their opportunism and antibiotic-resistance that lead to lung function decline (Courtney et al, 2004; Ren et al, 2012).

Prophages also contribute to the complex interplay of genetic and phenotypic diversity seen within the chronic lung as they enhance genetic variation through recombination events as they can encode for DNA recombinases (Lopes et al, 2010; Thorpe & Smith, 1998). This has also been shown in a CF related Pa population (Darch et al, 2015). Serine, tyrosine site specific recombinases and homologous recombinases have been identified in phages (Murphy, 2012). In *E. coli* the lambda Gam recombinase commercially known as lambda red has been shown to increase the rates of ssDNA recombination by ~ 1000 fold (Yu et al, 2003). Analogous genes to these recombinases are found in lambdoid-like dsDNA phages. These recombination events alongside mis-packaging or mis-excision of viral DNA leading to carriage of bacterial genes play a large role in how phages rapidly evolve after being able to infect the bacterial cell.

Hypotheses where phage-host interactions lead to variation in bacterial populations have been offered. One such proposition is “kill the winner”, it proposes that

the superior competitor (dominant bacteria) can sustain enough phage that can eventually reduce their host population size (Thingstad & Lignell, 1997). This is an example of where the phages help maintain host diversity through reducing the population size of the dominant bacteria. Another theory termed the “red queen” can be applied to the virus and their bacterial host, co-evolving constantly where neither absolutely eradicates the other however, drive each another to adapt and proliferate. Also there is the piggy back the winner hypothesis, looking at marine ecology and algal blooms it seems sensible and intuitive that if a phage is dormant in the dominant host then it can take advantage of the microbial population in the CF and BR patients lungs (Silveira & Rohwer, 2016).

Dynamic interplay between the bacteriophage and their host in the lung bacteria are shown to evolve in order to overcome phage infection (chapter 1.6.6). In contrast phage can overcome these strategies’ and simultaneously offer fitness through cell subversion to their hosts (chapter 1.6.7). The evolutionary dynamics are linked to bacterial survival in the chronic lung which leads to poor clinical outcomes for patients both for Pa (Hart & Winstanley, 2002) and Bcc (Ledson et al, 2002).

This thesis aimed to begin to study the role of temperate bacteriophages in chronic bacterial infection in CF and BR patients. Chapter 3 illustrates how the cross infection profiles of chemically induced temperate phages from 94 clinical Pa isolates consisting of 47 CF and 47 BR patients is different, but is linked to the originating disease state. The first and foremost finding presented in this study was that all the Pa isolates harbour at least one inducible temperate phage. This is supported by studies characterising bacteriophages of Pa in the lung (refs). The study aimed to relate the infection profiles of the induced bacteriophages and compare to the biology of phage and clinical data of the patient. The phage and their bacterial host range were therefore categorised based on disease progression as described in (chapter 2.5.1). The study presented in chapter 3 is therefore novel as it determines the infection profile of mixed Pa phage samples that were induced from a single clonal bacterial isolate. We

hypothesise that this would best represent phage induction in the lung from a single bacterium. It determines the potential of total virus interaction that links to the dynamic interplay between the host and bacteria. Pa bacteria have been well described to harbour multiple prophage regions (Winstanley et al, 2009) in their genomes (described in chapter 1.4). The study showed how the infection profile of the phage changed as the disease progresses and through PLS-DA modelling these phages could be grouped based on their infection profiles. We proposed that this could be due to the phage evolving strategies to evade bacterial host detection.

The study showed that phages induced from adult CF and > 10 BR Pa isolates had a higher infection rate when compared to their naïve counterpart's paediatric CF and < 10 BR Pa isolate. The study also showed that as the disease progressed, inducible phages could re-infect their originating bacterial host at a higher frequency (Tariq et al, 2015). This finding does not conform to the lambda phage infection model. The lambda repressor molecule binds to the superinfecting phages DNA operator preventing transcription. This study is not the first to describe superinfection of a lysogenic phage. However, this study illustrates the largest number of phages to show this trait experimentally in one study. Superinfection of isogenic phages and multiple infection sites have been previously described in *E. coli* with STX encoding phage (Allison et al, 2003; Fogg et al, 2007).

The thesis also aimed to characterise the phages at the genome level. The need to characterise bacteriophages as complete genomes can be exemplified by comparing the number of phage genome sequence available in GenBank. To date there are over 5600 viruses of which 2439 complete dsDNA bacteriophage genomes compared to in excess of 74000 prokaryotic genomes (NCBI, 2016). Bacteriophage genomes are a fraction of the size of their host genomes and at ~ 10^{31} phages with an estimated 10^{23} infections every second that can lead to a turnover rate of the entire phage population in just a few days (Suttle, 2007), fewer bacteriophage genomes are found on the GenBank database. This therefore shows the importance of the study in chapter 4 and the novelty

of the method to bioinformatically remove carry over bacterial contamination. This approach was novel as it aimed to characterise the mixed phage lysate population as described in chapter 3.

The key advantage of this metaviral method is we negate the use of plaque purification, phage proliferation and gradient isolation methods. Chapter 4 and 5 also aimed to annotate all the Pa and Bcc phages respectively, in the mixed lysate population. However, the study identified 2 key parameters that gave limitations to this approach. The first and foremost is that the sequence DNA yield which is crucial to assemble mixed phage populations and the second is highly conserved homology between extremely close phage genomes resulted in chimeric assemblies. The study showed how this was partially overcome by selecting the sequences based on the k-mer peaks and offered completion of phage genomes otherwise seen as chimeric assemblies. We were also able to increase contig sizes in the assemblies using the PriceTI software that is designed to complete or bridge gaps between sequence data by recalling paired end sequences with central gaps and building extra scaffolds within these assembly gaps. This helped resolve phage genomes once reasonable partial assemblies were constructed. Using this novel assembly approach in described in chapter 4, 107 Pa phages were presented (appendix 3), without PriceTI this would have been 80. The study showed that the most abundant phages present in the Pa isolates are F10-like phages. This study also identified that Phi297 and D3112 like phages were predominantly seen together in the mixed phage lysates. The class of the phages infecting Pa isolates were predominantly *Siphoviridae* class apart from H66 which was a *Podoviridae*. A novel highly conserved H66-like phage was identified 12 times in separate phage samples to exactly 61.772 kbp. The study hypothesised that their conservation could be linked to the disease as nine of the twelve were from BR Pa isolate.

Chapter 5 aimed to shows similar dynamics of phage and host relations between Bcc phage and their hosts. The study showed that all the Bcc isolates harboured at least one chemically inducible temperate phage. The study also showed that the phages

despite having genotypic variation exhibited similar infection rates. This was described for *multivoran* (n = 13 and *genomovar III A* (n = 22) isolates as they consisted the main cohort of the 47 Bcc. The study showed that Bcc phages could re-infect their originating host in 68% of the bacterial isolates. This finding was similar to the Pa samples and does not conform to the lambda repressor model of phage superinfection and should be studied further. The study went on to characterise these phages using the same approach as for the Pa sample. The key difference seen immediately between Bcc and Pa phage sequences was that they harboured fewer phages in their genomes. The Bcc phages were shown to be more diverse genetically and phenotypically. In chapter 5, 26 Bcc phages were presented which shared similarity to 13 different phages on the NCBI database compared to 107 Pa phages which shared some sequence homology to 10 phages, Taking into consideration that there are 133 Pa phage genomes compared to 31 Bcc phage genomes on GenBank database (NCBI, 2016). The study also showed that the Bcc phage genomes resembled six *Myoviridae*, four *Siphoviridae*, two *Podoviridae* and one unclassified type phages. Four BcepMu like phages were identified and shown to package a fragment of their host genomic DNA enhancing horizontal gene transfer. Phage DNA packaging by the BcepMu like phage increases transduction events and was hypothesised to drive evolution in the CF lung in chapter 5 (Summer et al, 2007). A study showed that the well characterised transducing Pa phages B3, F116L and G101 could infect Bcc isolates (Nzula et al, 2000). We did not test for cross species infection of the phages isolated. To a degree this study has shown intraspecies cross infection of temperate bacteriophage in Bcc sub-species. It might therefore, not be surprising to see multispecies infection of these phages.

A large study looking at 627 *Mycobacterium smegmatis* bacteriophage genomes, identified rich diversity, clustering in 28 distinct regions (Pope et al, 2015). The study identified these phages as having intra and inter-specific diversity with varying homology to different phages. The study went on to indicate that an inflow of genes from other sources were adding to the *Mycobacterium smegmatis* bacteriophage (Pope et al, 2015).

Similarly a study which looked at 337 phages infecting a range of *Enterobacteriaceae* (18 genera 31 species) showed their phages separate into 56 clusters primarily into temperate (24) and lytic (32) (Grose & Casjens, 2014). At this cluster level the phages were seen to exhibit > 50% sequence homology. The study went on to show that some clusters separated into more closely related sub-clusters of which 78 % were based on their host range (Grose & Casjens, 2014). Comparable phage genomic diversity was shown in a study looking at 98 *Bacillus* phages (Grose et al, 2014). These studies show that phages are understudied and underrepresented in the literature, often being overlooked, as the main focus of studies being their hosts. This thesis adds to the finding of these metagenomic studies and further emphasises temperate phage diversity within the chronic lung that is currently understated both for Pa and Bcc. In this study we do not see the depth of sequence diversity as described in the above studies however, we do show that at host specific level there are genetically diverse range of phages sharing little sequence homology to one another. This can be exemplified using the twelve H66-like phages. We showed that 11 of the 12 H66-like phage identified were similar to the last bp but they shared very little sequence similarity to the rest of the phages identified shown in the dot matrixes in chapter 4. This study will add draft genomes of these phages to the nucleotide database and therefore aid future comparisons of temperate Bcc and Pa phages.

Chapter 6 of this thesis looked to compare and contrast the BAM model (Barr et al, 2013b) previously described in lytic phages. This model illustrates phages that carry Ig-like domains on their tails and we proposed that temperate phages utilised these domains in order to bind to mucus, infect and propagate their bacterial hosts within the CF and BR lung. It was also proposed that lytic phages with this motif offered an extra layer of immunity against incoming bacteria that were sensitive to infection by these phages. This study hypothesised and illustrates that the incidence of Ig-like domains increased in Pa temperate phages in correlation to CF and BR disease progression. The study also incorporated an assembly comparison of VelvetOptimiser, IDBA-UD and

SPAdes. This study identified that SPAdes was the best at assembling phages in respect to identifying Ig-like domains. The study showed the assembly number of contigs of the SPAdes assembler to be lower and larger in comparison to IDBA-UD and VelvetOptimiser. A study showed that it can be difficult to get true assemblies of phages using *de novo* assembler and also recommended using a combination of the assemblies to get complete phage assemblies (Smits et al, 2014). The study showed that for the purpose of identifying Ig-like domains all three open source assemblers should be used as VelvetOptimiser identified an Ig-like domain that was not picked up in the assemblies of the SPAdes and IDBA-UD contigs (Tariq et al, 2015). Through the investigation of the Ig-like domains this chapter proposed an alternative hypothesis for temperate phage with respect to the BAM model. This study also proposed a domain architecture for two Ig-like domains; bacterial Ig-like domain group 2 (Big_2) and putative Ig domain (He_PIG).

The studies presented in this thesis helps characterise temperate phages in both Pa and Bcc bacterial isolate of CF and BR patients. The study showed although not for the first time that temperate Pa and Bcc bacteriophages have a varied infection profile and are in some cases capable of isogenic superinfection conforming to recent studies which suggest this enhances competition (Burns et al, 2015). A comparison between Pa and Bcc temperate phages that can be drawn from the thesis, is that the Pa bacteria were shown to harbour more phages with higher diversity compared to the Bcc phages isolates. We can postulate this from comparing the k-mer graphs and the number of phage seen in each sample. The polylysogenic nature of Pa may have had an effect of the sequence depth of each phage. The phage burst size, latent-period can be a contributor in the final phage titre and may not necessarily reflect the first phage induced out (Abedon et al, 2001). It is hypothesised that the difficulty in assembling Pa phage reflected the low sequence diversity or high sequence homology presented in this thesis. The relative ease of assembling Bcc phage could possibly be due to the fact the Bcc isolates harboured fewer phage, had phenotypical resemblance to *Myoviridae*,

Siphoviridae and *Podoviridae* type phages based on genetic similarity and shared little sequence homology.

In the CF Lung Bcc and Pa can coexist (Rogers et al, 2010) and a study showed that Bcc alter Pa through iron sequestration (Weaver & Kolter, 2004). The study showed that the siderophore, ornibactin produced by Bcc alters PA4467 gene expression. Ornibactin is not taken up by Pa and act by potentially limiting the iron available to Pa (Weaver & Kolter, 2004). Study have shown that in iron limiting environment Pa increase siderophore production, which correlates to the increase expression of the siderophore receptor on the cell surface (Cox & Adams, 1985). Siderophore production by Pa in the lung has been shown (Haas et al, 1991) and a link between the recent 'Ferrojan horse' hypothesis is paramount (Bonnain et al, 2016). The Ferrojan horse hypothesis postulates that T4 like phages harbour iron in their tails and compete for the siderophore receptors as a docking station and can therefore theoretically infect the cell. A study looking at biofilm in Pa and iron uptake showed that biofilms help accumulate and process large amount of iron compared to free living Pa (Lee & Beveridge, 2001). It would therefore not be surprising to see phages in an iron limiting lung environment take advantage of bacterial cells expressing siderophore receptors in order to infect. While concurrently phages of the Pa host range having shown potential to infect cross species into Bcc host. Coupled with expressing Ig-like domain in their tail fibres seen to increase with disease expression just exemplifies the complex dynamic of the lung driving phage and host evolution between two bacterial species and their phage. Figure 7.1 shows a flow diagram of the potential mechanisms used by Pa phage in order to adapt as survive in the chronic CF and BR lungs. Figure 7.1 encompasses our proposed BAM model for temperate phages (Tariq et al, 2015) and incorporates the piggy back the winner hypothesis (Silveira & Rohwer, 2016).

This thesis has shown that temperate bacteriophages to a degree have an involvement in Pa and Bcc bacterial evolution and may be involved in disease progression of the

lungs of CF and BR patients. Phage may amount to add a layer of intricacy that can ultimately have an effect on clinical outcomes of CF and BR patients.

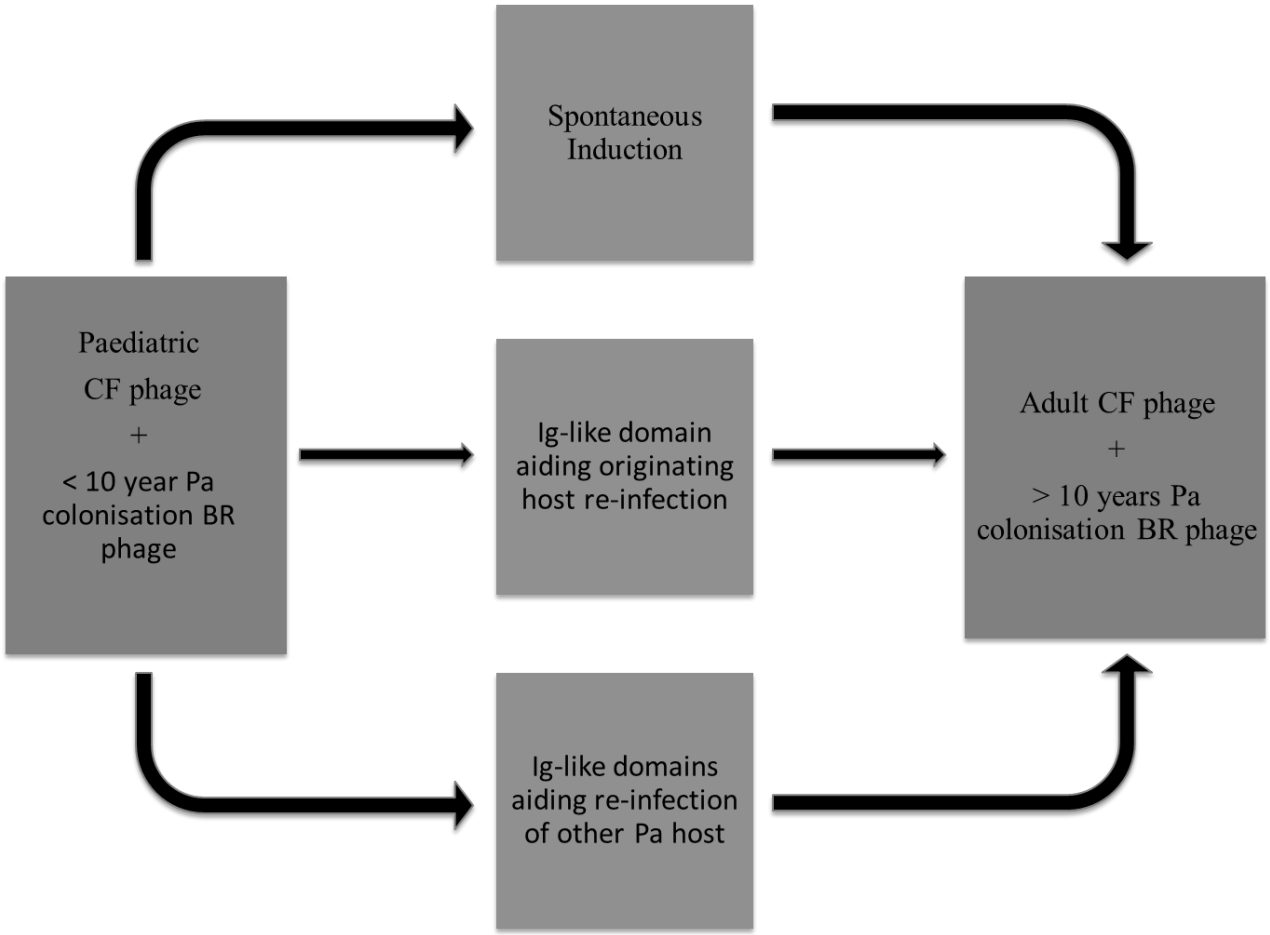


Figure 6.5 Possible strategies that Pa phage can evolve over time in the chronic lung. This flow diagram models the possible adaptation pathways in the chronic lung (Image adapted from (Tariq et al, 2015)).

7.2 Further Work

In order to fully elucidate if phage infection profiles in the cross infection studies both for Pa and Bcc are related to the disease progression it may be essential to first carry out a longitudinal study where a particular CF or BR patient is followed at a timely interval. A longitudinal study on the phages will also help answer the question if we see Pa phages genomic diversity due to disease progression and can help identify potential transduction events.

The phage genomes need validation to the correct orientation. The DNA in a phage virion is linear, but for different phages the sequence assembly may be linear or circular. The assembly of these phages is circular meaning the DNA molecules have terminal redundancy and can be difficult to establish true cos sites. Therefore the phage will be presented as in phage T4 and therefore it would be necessary to break it at an arbitrary position in order to get a linear sequence to submit to GenBank. This break is at the sequence just upstream from the terminase gene. However, this is only possible for phage with predicted terminase. Another method could be to break the sequence in the centre and make them the end. Mapping the reads back to the sequence would theoretically identify a true linear phage ends. The genome assembly still requires manual curation where the predicted genes are pipeline through RAST part of the Phantome tools. These annotations use Glimmer and GeneMark to call and predict open reading frames and may offer improved and accurate predictions.

The samples that failed to generate enough sequence yields or had chimeric assemblies need to be revisited. Using meta assemblers designed to assemble meta assemblies may resolve more phages that failed to assemble.

It would be informative to test cross species infectivity of these phages and whether they are able to infect Pa and Bcc isolates. Correlated back to the Pa disease aetiologies and to see if Pa phages with disease progression lose their broad range host infectivity. A cross species infection profile would test this hypothesis to a greater degree.

8. Reference

Abedon S (2011) Phage therapy pharmacology: calculating phage dosing. *Advanced Applied Microbiology* **77:** 1-40

Abedon ST, Herschler TD, Stopar D (2001) Bacteriophage latent-period evolution as a response to resource availability. *Applied and Environmental Microbiology* **67:** 4233-4241

Adams MH (1959) *Bacteriophages*, New York,: Interscience Publishers.

Aksyuk AA, Rossmann MG (2011) Bacteriophage assembly. *Viruses* **3:** 172-203

Allison HE (2007) Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens. *Future Microbiology* **2:** 165-174

Allison HE, Sergeant MJ, James CE, Saunders JR, Smith DL, Sharp RJ, Marks TS, McCarthy AJ (2003) Immunity profiles of wild-type and recombinant shiga-like toxin-encoding bacteriophages and characterization of novel double lysogens. *Infection and Immunity* **71:** 3409-3418

Altuvia S, Locker-Giladi H, Koby S, Ben-Nun O, Oppenheim AB (1987) RNase III stimulates the translation of the *cIII* gene of bacteriophage lambda. *Proceedings of the National Academy of Sciences of the United States of America* **84:** 6511-6515

Ames GF, Mimura CS, Shyamala V (1990) Bacterial periplasmic permeases belong to a family of transport proteins operating from *Escherichia coli* to human: Traffic ATPases. *FEMS Microbiology Reviews* **6:** 429-446

Andreolli M, Lampis S, Zenaro E, Salkinoja-Salonen M, Vallini G (2011) *Burkholderia fungorum* DBT1: a promising bacterial strain for bioremediation of PAHs-contaminated soils. *FEMS Microbiology Letters* **319:** 11-18

Andres D, Hanke C, Baxa U, Seul A, Barbirz S, Seckler R (2010) Tailspike Interactions with Lipopolysaccharide Effect DNA Ejection from Phage P22 Particles in Vitro. *Journal of Biological Chemistry* **285:** 36768-36775

Angus AA, Agapakis CM, Fong S, Yerrapragada S, Estrada-de los Santos P, Yang P, Song N, Kano S, Caballero-Mellado J, de Faria SM, Dakora FD, Weinstock G, Hirsch AM (2014) Plant-Associated Symbiotic *Burkholderia* Species Lack Hallmark Strategies Required in Mammalian Pathogenesis. *PloS One* **9**

Arber W (1979) Promotion and Limitation of Genetic Exchange. *Science* **205:** 361-365

Aris RM, Routh JC, LiPuma JJ, Heath DG, Gilligan PH (2001) Lung transplantation for cystic fibrosis patients with *Burkholderia cepacia* complex - Survival linked to genomovar type. *American Journal of Respiratory and Critical Care Medicine* **164:** 2102-2106

Athanazio R (2012) Airway disease: similarities and differences between asthma, COPD and bronchiectasis. *Clinics* **67:** 1335-1343

Athanazio RA, Rached SZ, Rohde C, Pinto RC, Fernandes FLA, Stelmach R (2010) Should the bronchiectasis treatment given to cystic fibrosis patients be extrapolated to those with bronchiectasis from other causes? *Jornal Brasileiro De Pneumologia* **36:** 425-431

Baldan R, Cigana C, Testa F, Bianconi I, De Simone M, Pellin D, Di Serio C, Bragonzi A, Cirillo DM (2014) Adaptation of Pseudomonas aeruginosa in Cystic Fibrosis Airways Influences Virulence of *Staphylococcus aureus* In Vitro and Murine Models of Co-Infection. *PloS One* **9**

Bandyopadhyay PK, Studier FW, Hamilton DL, Yuan R (1985) Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *Journal of Molecular Biology* **182:** 567-578

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19:** 455-477

Barnhart BJ, Cox SH, Jett JH (1976) Prophage Induction and Inactivation by UV Light. *Journal of Virology* **18:** 950-955

Barondess JJ, Beckwith J (1995) Bor Gene of Phage-Lambda, Involved in Serum Resistance, Encodes a Widely Conserved Outer-Membrane Lipoprotein. *Journal of Bacteriology* **177:** 1247-1253

Barr JJ, Auro R, Furlan M, Whiteson KL, Erb ML, Pogliano J, Stotland A, Wolkowicz R, Cutting AS, Doran KS (2013a) Bacteriophage adhering to mucus provide a non–host-derived immunity. *Proceedings of the National Academy of Sciences* **110:** 10771-10776

Barr JJ, Auro R, Furlan M, Whiteson KL, Erb ML, Pogliano J, Stotland A, Wolkowicz R, Cutting AS, Doran KS, Salamon P, Youle M, Rohwer F (2013b) Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 10771-10776

Bartell PF, Orr TE, Reese JF, Imaeda T (1971) Interaction of *Pseudomonas* Bacteriophage-2 with Slime Polysaccharide and Lipopolysaccharide of *Pseudomonas aeruginosa* Strain Bi. *Journal of Virology* **8:** 311-&

Bergh O, Borsheim KY, Bratbak G, Heldal M (1989) High Abundance of Viruses Found in Aquatic Environments. *Nature* **340:** 467-468

Berkane E, Orlik F, Stegmeier JF, Charbit A, Winterhalter M, Benz R (2006) Interaction of bacteriophage Lambda with its cell surface receptor: An in vitro study of binding of the viral tail protein gpJ to LamB (Maltoporin). *Biochemistry* **45:** 2708-2720

Bernard P, Couturier M (1992) Cell Killing by the F-Plasmid Ccdb Protein Involves Poisoning of DNA-Topoisomerase-Ii Complexes. *Journal of Molecular Biology* **226:** 735-745

Bittar F, Richet H, Dubus JC, Reynaud-Gaubert M, Stremler N, Sarles J, Raoult D, Rolain JM (2008) Molecular Detection of Multiple Emerging Pathogens in Sputa from Cystic Fibrosis Patients. *PloS One* **3**

Black LW (1989) DNA Packaging in dsDNA Bacteriophages. *Annual Review of Microbiology* **43:** 267-292

Black LW, Rao VB (2012) Structure, Assembly, and DNA Packaging of the Bacteriophage T4 Head. *Advances in Virus Research, Vol 82: Bacteriophages, Pt A* **82:** 119-153

Blackburn L, Brownlee K, Conway S, Denton M (2004) 'Cepacia syndrome' with *Burkholderia multivorans*, 9 years after initial colonization. *Journal of Cystic Fibrosis* **3:** 133-134

Bonadia LC, Marson FAD, Ribeiro JD, Paschoal IA, Pereira MC, Ribeiro AF, Bertuzzo CS (2014) CFTR genotype and clinical outcomes of adult patients carried as cystic fibrosis disease. *Gene* **540:** 183-190

Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493:** 429-432

Bonnain C, Breitbart M, Buck KN (2016) The Ferrojan Horse Hypothesis: Iron-Virus Interactions in the Ocean. *Frontiers in Marine Science* **3**

Borowitz D (2005) Update on the evaluation of pancreatic exocrine status in cystic fibrosis. *Current Opinion in Pulmonary Medicine* **11:** 524-527

Boyton RJ, Altmann DM (2016) Bronchiectasis: Current Concepts in Pathogenesis, Immunology, and Microbiology. *Annual Review of Pathology: Mechanisms of Disease, Vol 11* **11:** 523-554

Braid MD, Silhavy JL, Kitts CL, Cano RJ, Howe MM (2004) Complete genomic sequence of bacteriophage B3, a Mu-like phage of *Pseudomonas aeruginosa*. *Journal of Bacteriology* **186:** 6560-6574

Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin EV, van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321:** 960-964

Brussow H (2005) Phage therapy: the *Escherichia coli* experience. *Microbiology* **151:** 2133-2140

Burkholder WH (1950) Sour skin, a bacterial rot of onion bulbs. *Phytopathology* **40**

Burns N, James CE, Harrison E (2015) Polylysogeny magnifies competitiveness of a bacterial pathogen in vivo. *Evolutionary Applications* **8:** 346-351

Byrne M, Kropinski AM (2005) The genome of the *Pseudomonas aeruginosa* generalized transducing bacteriophage F116. *Gene* **346:** 187-194

Caballero JD, Clark ST, Coburn B, Zhang Y, Wang PW, Donaldson SL, Tullis DE, Yau YCW, Waters VJ, Hwang DM, Guttman DS (2015) Selective Sweeps and Parallel Pathoadaptation Drive *Pseudomonas aeruginosa* Evolution in the Cystic Fibrosis Lung. *mBio* **6**

Cady KC, Bondy-Denomy J, Heussler GE, Davidson AR, O'Toole GA (2012) The CRISPR/Cas Adaptive Immune System of *Pseudomonas aeruginosa* Mediates Resistance to Naturally Occurring and Engineered Phages. *Journal of Bacteriology* **194:** 5728-5738

Callaghan M, McClean S (2012) Bacterial host interactions in cystic fibrosis. *Current Opinion in Microbiology* **15:** 71-77

Capaldo FN, Barbour SD (1975) The role of the *rec* genes in the viability of *Escherichia coli* K12. *Basic Life Sciences* **5A:** 405-418

Carlier A, Agnoli K, Pessi G, Suppiger A, Jenul C, Schmid N, Tummler B, Pinto-Carbo M, Eberl L (2014) Genome Sequence of *Burkholderia cenocepacia* H111, a Cystic Fibrosis Airway Isolate. *Genome Announcements* **2**

Carmody LA, Gill JJ, Summer EJ, Sajjan US, Gonzalez CF, Young RF, LiPuma JJ (2010) Efficacy of Bacteriophage Therapy in a Model of *Burkholderia cenocepacia* Pulmonary Infection. *Journal of Infectious Diseases* **201:** 264-271

Casjens S, Adams MB, Hall C, King J (1985) Assembly-Controlled Autogenous Modulation of Bacteriophage-P22 Scaffolding Protein Gene-Expression. *Journal of Virology* **53:** 174-179

Casjens S, Hendrix R (1974) Comments on the arrangement of the morphogenetic genes of bacteriophage lambda. *Journal of Molecular Biology* **90:** 20-25

Casjens SR, Hendrix RW (2015) Bacteriophage lambda: Early pioneer and still relevant. *Virology* **479:** 310-330

Caspar DL, Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposia on Quantitative Biology* **27:** 1-24

Castellani C, Cuppens H, Macek M, Cassinian JJ, Kerern E, Durie P, Tullis E, Assael BM, Bombieri C, Brown A, Casals T, Claustres M, Cutting GR, Dequeker E, Dodge J, Doull I, Farrell P, Ferec C, Girodon E, Johannesson M, Kerem B, Knowles M, Munck A, Pignatti PF, Radojkovic D, Rizzotti P, Schwarz M, Stuhnnann M, Tzetis M, Zielenski J, Elborn JS (2008) Consensus on the use and interpretation of cystic fibrosis mutation analysis in clinical practice. *Journal of Cystic Fibrosis* **7:** 179-196

Castillo FJ, Bartell PF (1976) Localization and functional role of the *Pseudomonas* bacteriophage 2 depolymerase. *Journal of Virology* **18:** 701-708

Cazares A, Mendoza-Hernandez G, Guarneros G (2014) Core and accessory genome architecture in a group of *Pseudomonas aeruginosa* Mu-like phages. *BMC Genomics* **15**

Ceyssens PJ, Glonti T, Kropinski NM, Lavigne R, Chanishvili N, Kulakov L, Lashkhi N, Tediashvili M, Merabishvili M (2011) Phenotypic and genotypic variations within a single bacteriophage species. *Virology Journal* **8**

Ceyssens PJ, Lavigne R (2010) Bacteriophages of *Pseudomonas*. *Future Microbiology* **5:** 1041-1055

Chain E, Florey HW, Adelaide MB, Gardner AD, Heatley NG, Jennings MA, Orr-Ewing J, Sanders AG, Peltier LF (2005) The Classic - Penicillin as a chemotherapeutic agent (Reprinted from Lancet, vol 24, pg 226-231, 1940). *Clinical Orthopaedics and Related Research***:** 23-26

Cheng K, Smyth RL, Govan JRW, Doherty C, Winstanley C, Denning N, Heaf DP, vanSaene H, Hart CA (1996) Spread of beta-lactam-resistant *Pseudomonas aeruginosa* in a cystic fibrosis clinic. *Lancet* **348:** 639-642

Chhibber S, Kaur T, Kaur S (2013) Co-Therapy Using Lytic Bacteriophage and Linezolid: Effective Treatment in Eliminating Methicillin Resistant *Staphylococcus aureus* (MRSA) from Diabetic Foot Infections. *PloS One* **8**

Chibani-Chennoufi S, Bruttin A, Dillmann ML, Brussow H (2004) Phage-host interaction: an ecological perspective. *Journal of Bacteriology* **186:** 3677-3686

Chibeu A, Ceyssens PJ, Hertveldt K, Volckaert G, Cornelis P, Matthijs S, Lavigne R (2009) The adsorption of *Pseudomonas aeruginosa* bacteriophage phiKMV is dependent on expression regulation of type IV pili genes. *FEMS Microbiology Letters* **296:** 210-218

Chierakul W, Anunnatsiri S, Short JM, Maharjan B, Mootsikapun P, Simpson AJH, Limmathurotsakul D, Cheng AC, Stepniewska K, Newton PN, Chaowagul W, White NJ, Peacock SJ, Day NP, Chetchotisakd P (2005) Two randomized controlled trials of ceftazidime alone versus ceftazidime in combination with trimethoprim-sulfamethoxazole for the treatment of severe melioidosis. *Clinical Infectious Diseases* **41:** 1105-1113

Chopin MC, Chopin A, Bidnenko E (2005) Phage abortive infection in lactococci: variations on a theme. *Current Opinion in Microbiology* **8:** 473-479

Chu CS, Trapnell BC, Curristin S, Cutting GR, Crystal RG (1993) Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. *Nature Genetics* **3:** 151-156

Chung IY, Jang HJ, Bae HW, Cho YH (2014) A phage protein that inhibits the bacterial ATPase required for type IV pilus assembly. *Proceedings of the National Academy of Sciences of the United States of America* **111:** 11503-11508

Cihlar RL, Lessie TG, Holt SC (1978) Characterization of bacteriophage CP1, an organic solvent sensitive phage associated with *Pseudomonas cepacia*. *Canadian Journal of Microbiology* **24:** 1404-1412

Coenye T, Vandamme P (2003) Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environmental Microbiology* **5:** 719-729

Coenye T, Vandamme P, Govan JR, LiPuma JJ (2001) Taxonomy and identification of the *Burkholderia cepacia* complex. *Journal of Clinical Microbiology* **39:** 3427-3436

Cohen-Cymberknoh M, Shoseyov D, Kerem E (2011) Managing Cystic Fibrosis Strategies That Increase Life Expectancy and Improve Quality of Life. *American Journal of Respiratory and Critical Care Medicine* **183:** 1463-1471

Collins FS, Morgan M, Patrinos A (2003) The human genome project: Lessons from large-scale biology. *Science* **300:** 286-290

Comolli JC, Hauser AR, Waite L, Whitchurch CB, Mattick JS, Engel JN (1999) *Pseudomonas aeruginosa* gene products PilT and PilU are required for cytotoxicity in vitro and virulence in a mouse model of acute pneumonia. *Infection and Immunity* **67:** 3625-3630

Cone RA (2009) Barrier properties of mucus. *Advanced Drug Delivery Reviews* **61:** 75-85

Consortium TCFGA (1994) Population variation of common cystic fibrosis mutations. The Cystic Fibrosis Genetic Analysis Consortium. *Human Mutation* **4:** 167-177

Costerton JW, Stewart PS, Greenberg EP (1999) Bacterial biofilms: a common cause of persistent infections. *Science* **284:** 1318-1322

Courtney JM, Dunbar KE, McDowell A, Moore JE, Warke TJ, Stevenson M, Elborn JS (2004) Clinical outcome of *Burkholderia cepacia* complex infection in cystic fibrosis adults. *Journal of Cystic Fibrosis* **3:** 93-98

Cox CD (1986) Role of pyocyanin in the acquisition of iron from transferrin. *Infection and Immunity* **52:** 263-270

Cox CD, Adams P (1985) Siderophore Activity of Pyoverdin for *Pseudomonas aeruginosa*. *Infection and Immunity* **48:** 130-138

Cox MJ, Allgaier M, Taylor B, Baek MS, Huang YJ, Daly RA, Karaoz U, Andersen GL, Brown R, Fujimura KE, Wu B, Tran D, Koff J, Kleinhenz ME, Nielson D, Brodie EL, Lynch SV (2010) Airway Microbiota and Pathogen Abundance in Age-Stratified Cystic Fibrosis Patients. *PloS One* **5**

CRISPRdb. (2016) CRISPRs web server. Vol. 2016.

Cystic Fibrosis Foundation. (2014) Cystic Fibrosis Foundation. Vol. 2014.

Cystic fibrosis foundation. (2016) Cystic Fibrosis Foundation. Vol. 2016.

Cystic Fibrosis mutation database. (2016) Cystic Fibrosis mutation database. Vol. 20156.

d'Herelle F (1917) Sur un microbe invisible antagonistic des bacilles dysenterique. *Comptes Rendus de l'Académie des Sciences Série III: Sciences de la Vie* **165:** 373-375

Dajani AS (1972) The scalded-skin syndrome: relation to phage-group II staphylococci. *The Journal of Infectious Disease* **125:** 548-551

Daniels T (2010) Physiotherapeutic management strategies for the treatment of cystic fibrosis in adults. *The Journal of Multidisciplinary Healthcare* **3:** 201-212

Darch SE, McNally A, Harrison F, Corander J, Barr HL, Paszkiewicz K, Holden S, Fogarty A, Crusz SA, Diggle SP (2015) Recombination is a key driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic fibrosis infection. *Scientific Reports* **5**

Datta DB, Arden B, Henning U (1977) Major Proteins of *Escherichia coli* Outer Cell-Envelope Membrane as Bacteriophage Receptors. *Journal of Bacteriology* **131:** 821-829

Davies EV, James CE, Williams D, O'Brien S, Fothergill JL, Haldenby S, Paterson S, Winstanley C, Brockhurst MA (2016) Temperate phages both mediate and drive adaptive evolution in pathogen biofilms. *Proceedings of the National Academy of Sciences of the United States of America* **113:** 8266-8271

Davis PB, Yasothan U, Kirkpatrick P (2012) Ivacaftor. *Nature Reviews Drug Discovery* **11:** 349-350

De Boeck K, Malfroot A, Van Schil L, Lebecque P, Knoop C, Govan JRW, Doherty C, Laevens S, Vandamme P, Burkholderia BotB (2004) Epidemiology of *Burkholderia cepacia* complex colonisation in cystic fibrosis patients. *European Respiratory Journal* **23:** 851-856

De Soyza A, McDowell A, Archer L, Dark JH, Elborn SJ, Mahenthiralingam E, Gould K, Corris PA (2001) *Burkholderia cepacia* complex genomovars and pulmonary transplantation outcomes in patients with cystic fibrosis. *Lancet* **358:** 1780-1781

De Wyngaert M, Hinkle DC (1979) Involvement of DNA gyrase in replication and transcription of bacteriophage T7 DNA. *Journal of Virology* **29:** 529-535

Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23:** 673-679

Delhaes L, Monchy S, Frealle E, Hubans C, Salleron J, Leroy S, Prevotat A, Wallet F, Wallaert B, Dei-Cas E, Sime-Ngando T, Chabe M, Viscogliosi E (2012) The Airway Microbiota in Cystic Fibrosis: A Complex Fungal and Bacterial Community-Implications for Therapeutic Management. *PloS One* **7**

DeShazer D (2004) Genomic diversity of *Burkholderia pseudomallei* clinical isolates: Subtractive hybridization reveals a *Burkholderia mallei* - Specific prophage in *B. pseudomallei* 1026b. *Journal of Bacteriology* **186:** 3938-3950

Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology* **190:** 1390-1400

Dietrich LEP, Price-Whelan A, Petersen A, Whiteley M, Newman DK (2006) The phenazine pyocyanin is a terminal signalling factor in the quorum sensing network of *Pseudomonas aeruginosa*. *Molecular Microbiology* **61:** 1308-1321

Dillingham MS, Kowalczykowski SC (2008) RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. *Microbiology and Molecular Biology Reviews* **72:** 642-+

Dodd IB, Perkins AJ, Tsemitsidis D, Egan JB (2001) Octamerization of lambda CI repressor is needed for effective repression of P-RM and efficient switching from lysogeny. *Genes and Development* **15:** 3013-3022

Drulis-Kawa Z, Majkowska-Skrobek G, Maciejewska B (2015) Bacteriophages and Phage-Derived Proteins - Application Approaches. *Current Medicinal Chemistry* **22:** 1757-1773

Drulis-Kawa Z, Olszak T, Danis K, Majkowska-Skrobek G, Ackermann HW (2014) A giant *Pseudomonas* phage from Poland. *Archives of Virology* **159:** 567-572

Duff RM, Simmonds NJ, Davies JC, Wilson R, Alton EW, Pantelidis P, Cox MJ, Cookson WOCM, Bilton D, Moffatt MF (2013) A molecular comparison of microbial communities in bronchiectasis and cystic fibrosis. *European Respiratory Journal* **41:** 991-993

Dugueperoux I, De Braekeleer M (2005) The CFTR 3849+10kbC->T and 2789+5G->A alleles are associated with a mild CF phenotype. *European Respiratory Journal* **25:** 468-473

Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GG, Boling L, Barr JJ, Speth DR, Seguritan V, Aziz RK (2014) A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nature communications* **5**

Dy RL, Przybilski R, Semeijn K, Salmond GPC, Fineran PC (2014) A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic Acids Research* **42:** 4590-4605

Earnshaw WC, King J, Harrison SC, Eiserling FA (1978) Structural Organization of DNA Packaged within Heads of T4 Wild-Type, Isometric and Giant Bacteriophages. *Cell* **14:** 559-568

Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit MA (2016) Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *The ISME Journal*

Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, Young VB, Toews GB, Curtis JL, Sundaram B, Martinez FJ, Huffnagle GB (2011) Analysis of the Lung Microbiome in the "Healthy" Smoker and in COPD. *PloS One* **6**

Estrada-de los Santos P, Vinuesa P, Martinez-Aguilar L, Hirsch AM, Caballero-Mellado J (2013) Phylogenetic Analysis of *Burkholderia* Species by Multilocus Sequence Analysis. *Current Microbiology* **67:** 51-60

Evilevitch A, Lavelle L, Knobler CM, Raspaud E, Gelbart WM (2003) Osmotic pressure inhibition of DNA ejection from phage. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 9292-9295

Farber S, Shwachman H, Maddock CL (1943) Pancreatic Function and Disease in Early Life. I. Pancreatic Enzyme Activity and the Celiac Syndrome. *The Journal of clinical investigation* **22:** 827-838

Fauquet C.M., Schrock J.R. (2006) Virus Classification. *Classification of Viruses*. ESU printing services, Emporia State Univesity, Vol. 2016.

Ferec C, Cutting GR (2012) Assessing the Disease-Liability of Mutations in CFTR. *Cold Spring Harbor Perspectives in Medicine* **2**

Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GP (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair.

Proceedings of the National Academy of Sciences of the United States of America **106:** 894-899

Fleming A (2001) On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. *Bulletin of the World Health Organization* **79:** 780-790

Fodor AA, Klem ER, Gilpin DF, Elborn JS, Boucher RC, Tunney MM, Wolfgang MC (2012) The Adult Cystic Fibrosis Airway Microbiota Is Stable over Time and Infection Type, and Highly Resilient to Antibiotic Treatment of Exacerbations. *PloS One* **7**

Fogg PC, Saunders JR, McCarthy AJ, Allison HE (2012) Cumulative effect of prophage burden on Shiga toxin production in *Escherichia coli*. *Microbiology* **158:** 488-497

Fogg PCM, Gossage SM, Smith DL, Saunders JR, McCarthy AJ, Allison HE (2007) Identification of multiple integration sites for Stx-phage Phi 24(B) in the *Escherichia coli* genome, description of a novel integrase and evidence for a functional anti-repressor. *Microbiology-Sgm* **153:** 4098-4110

Fogg PCM, Rigden DJ, Saunders JR, McCarthy AJ, Allison HE (2011) Characterization of the relationship between integrase, excisionase and antirepressor activities associated with a superinfecting Shiga toxin encoding bacteriophage. *Nucleic Acids Research* **39:** 2116-2129

Fozo EM, Hemm MR, Storz G (2008) Small Toxic Proteins and the Antisense RNAs That Repress Them. *Microbiology and Molecular Biology Reviews* **72:** 579-589

Fraser JS, Maxwell KL, Davidson AR (2007) Immunoglobulin-like domains on bacteriophage: weapons of modest damage? *Current Opinion in Microbiology* **10:** 382-387

Fraser JS, Yu Z, Maxwell KL, Davidson AR (2006) Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *J Mol Biol* **359:** 496-507

Garvey P, Fitzgerald GF, Hill C (1995) Cloning and DNA sequence analysis of two abortive infection phage resistance determinants from the lactococcal plasmid pNP40. *Applied and Environmental Microbiology* **61:** 4321-4328

Gasparini P, Nunes V, Savoia A, Dognini M, Morral N, Gaona A, Bonizzato A, Chillon M, Sangiuolo F, Novelli G, et al. (1991) The search for south European cystic fibrosis mutations: identification of two new mutations, four variants, and intronic sequences. *Genomics* **10:** 193-200

Geborek A, Hjelte L (2011) Association between genotype and pulmonary phenotype in cystic fibrosis patients with severe mutations. *Journal of Cystic Fibrosis* **10:** 187-192

Gessard C (1984) On the Blue and Green Coloration That Appears on Bandages. *Reviews of Infectious Diseases* **6:** S775-S776

Gibson RL, Burns JL, Ramsey BW (2003) Pathophysiology and management of pulmonary infections in cystic fibrosis. *American Journal of Respiratory and Critical Care Medicine* **168:** 918-951

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17:** 333-351

Gorter FA, Hall AR, Buckling A, Scanlan PD (2015) Parasite host range and the evolution of host resistance. *Journal of Evolutionary Biology* **28:** 1119-1130

Greiling D, Clark RAF (1997) Fibronectin provides a conduit for fibroblast transmigration from collagenous stroma into fibrin clot provisional matrix. *Journal of Cell Science* **110:** 861-870

Grose JH, Casjens SR (2014) Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family *Enterobacteriaceae*. *Virology* **468:** 421-443

Grose JH, Jensen GL, Burnett SH, Breakwell DP (2014) Genomic comparison of 93 *Bacillus* phages reveals 12 clusters, 14 singletons and remarkable diversity (vol 15, 855, 2014). *BMC Genomics* **15**

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29:** 1072-1075

Guss AM, Roeselers G, Newton ILG, Young CR, Klepac-Ceraj V, Lory S, Cavanaugh CM (2011) Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *The ISME Journal* **5:** 20-29

Haas B, Kraut J, Marks J, Zanker SC, Castignetti D (1991) Siderophore Presence in Sputa of Cystic-Fibrosis Patients. *Infection and Immunity* **59:** 3997-4000

Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP (2009) RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* **139:** 945-956

Hansson GC (2012) Role of mucus layers in gut infection and inflammation. *Current Opinion in Microbiology* **15:** 57-62

Hantke K (1978) Major outer membrane proteins of *E. coli* K12 serve as receptors for the phages T2 (protein Ia) and 434 (protein Ib). *Molecular and General Genetics* **164:** 131-135

Hart CA, Winstanley C (2002) Persistent and aggressive bacteria in the lungs of cystic fibrosis children. *British Medical Bulletin* **61:** 81-96

Hashemolhosseini S, Montag D, Kramer L, Henning U (1994) Determinants of receptor specificity of coliphages of the T4 family. A chaperone alters the host range. *Journal of Molecular Biology* **241:** 524-533

Hatfull GF (2008) Bacteriophage genomics. *Current Opinion in Microbiology* **11:** 447-453

Hauser AR, Jain M, Bar-Meir M, McColley SA (2011) Clinical Significance of Microbial Infection and Adaptation in Cystic Fibrosis. *Clinical Microbiology Reviews* **24:** 29-70

Haussler S, Ziegler I, Lottel A, von Gotz F, Rohde M, Wehmhohner D, Saravanamuthu S, Tummler B, Steinmetz I (2003) Highly adherent small-colony variants of *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Journal of Medical Microbiology* **52:** 295-301

Hayashi T, Baba T, Matsumoto H, Terawaki Y (1990) Phage Conversion of Cytotoxin Production in *Pseudomonas aeruginosa*. *Molecular Microbiology* **4:** 1703-1709

Hayashi T, Kamio Y, Hishinuma F, Usami Y, Titani K, Terawaki Y (1989) *Pseudomonas aeruginosa* cytotoxin: the nucleotide sequence of the gene and the mechanism of activation of the protoxin. *Molecular Microbiology* **3:** 861-868

Hendrix RW (2002) Bacteriophages: evolution of the majority. *Theoretical Population Biology* **61:** 471-480

Hendrix RW (2003) Bacteriophage genomics. *Current Opinion in Microbiology* **6:** 506-511

Herskowitz I, Hagen D (1980) The Lysis Lysogeny Decision of Phage Lambda - Explicit Programming and Responsiveness. *Annual Review of Genetics* **14:** 399-445

Higenbottam T, Cochrane GM, Clark TJH, Turner D, Millis R, Seymour W (1980) Bronchial Disease in Ulcerative-Colitis. *Thorax* **35:** 581-585

Hoiby N, Ciofu O, Bjarnsholt T (2010) *Pseudomonas aeruginosa* biofilms in cystic fibrosis. *Future Microbiology* **5:** 1663-1674

Holden MT, Titball RW, Peacock SJ, Cerdeno-Tarraga AM, Atkins T, Crossman LC, Pitt T, Churcher C, Mungall K, Bentley SD, Sebaihia M, Thomson NR, Bason N, Beacham IR, Brooks K, Brown KA, Brown NF, Challis GL, Cherevach I, Chillingworth T, Cronin A, Crossett B, Davis P, DeShazer D, Feltwell T, Fraser A, Hance Z, Hauser H, Holroyd S, Jagels K, Keith KE, Maddison M, Moule S, Price C, Quail MA, Rabbinowitsch E, Rutherford K, Sanders M, Simmonds M, Songsivilai S, Stevens K, Tumapa S, Vesaratchavest M, Whitehead S, Yeats C, Barrell BG, Oyston PC, Parkhill J (2004) Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei.

Proceedings of the National Academy of Sciences of the United States of America **101:** 14240-14245

Holden MTG, Seth-Smith HMB, Crossman LC, Sebaihia M, Bentley SD, Cerdeno-Tarraga AM, Thomson NR, Bason N, Quail MA, Sharp S, Cherevach I, Churcher C, Goodhead I, Hauser H, Holroyd N, Mungall K, Scott P, Walker D, White B, Rose H, Iversen P, Mil-Homens D, Rocha EPC, Fialho AM, Baldwin A, Dowson C, Barrell BG, Govan JR, Vandamme P, Hart CA, Mahenthiralingam E, Parkhill J (2009) The Genome of *Burkholderia cenocepacia* J2315, an Epidemic Pathogen of Cystic Fibrosis Patients (vol 191, pg 261, 2009). *Journal of Bacteriology* **191:** 2907-2907

Holmes RK (2000) Biology and molecular epidemiology of diphtheria toxin and the tox gene. *Journal of Infectious Diseases* **181:** S156-S167

Homolog.us. (2016) An Explanation of Velvet Parameter exp_cov. Vol. 2016.

Housby JN, Mann NH (2009) Phage therapy. *Drug Discovery Today* **14:** 536-540

Hubner A, Danganan CE, Xun LY, Chakrabarty AM, Hendrickson W (1998) Genes for 2,4,5-trichlorophenoxyacetic acid metabolism in *Burkholderia cepacia* AC1100: Characterization of the tftC and tftD genes and locations of the tft operons on multiple replicons. *Applied and Environmental Microbiology* **64:** 2086-2093

Hughes DT, Clarke MB, Yamamoto K, Rasko DA, Sperandio V (2009) The QseC Adrenergic Signaling Cascade in Enterohemorrhagic *E. coli* (EHEC). *PLoS Pathogens* **5**

Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**

ICTV. (2016) International Committee on Taxonomy of Viruses

Idury RM, Waterman MS (1995) A new algorithm for DNA sequence assembly. *Journal of Computational Biology* **2:** 291-306

Illumina. (2016) paired end sequencing. Vol. 2016.

James CE, Davies EV, Fothergill JL, Walshaw MJ, Beale CM, Brockhurst MA, Winstanley C (2015) Lytic activity by temperate phages of *Pseudomonas aeruginosa* in long-term cystic fibrosis chronic lung infections. *The ISME Journal* **9:** 1391-1398

James CE, Fothergill JL, Kalwij H, Hall AJ, Cottell J, Brockhurst MA, Winstanley C (2012) Differential infection properties of three inducible prophages from an epidemic strain of *Pseudomonas aeruginosa*. *BMC Microbiology* **12**

Jansen R, Embden JD, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* **43:** 1565-1575

Johnson A, Meyer BJ, Ptashne M (1978) Mechanism of action of the cro protein of bacteriophage lambda. *Proceedings of the National Academy of Sciences of the United States of America* **75:** 1783-1787

Kaiser AD, Jacob F (1957) Recombination between related temperate bacteriophages and the genetic control of immunity and prophage localization. *Virology* **4:** 509-521

Kalish LA, Waltz DA, Dovey M, Potter-Bynoe G, McAdam AJ, LiPuma JJ, Gerard C, Goldmann D (2006) Impact of *Burkholderia dolosa* on lung function and survival in cystic fibrosis. *American Journal of Respiratory and Critical Care Medicine* **173:** 421-425

Karaolis DKR, Somara S, Maneval DR, Johnson JA, Kaper JB (1999) A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399:** 375-379

Katsura I (1981) Structure and function of the major tail protein of bacteriophage lambda: Mutants having small major tail protein molecules in their virion. *Journal of Molecular Biology* **146:** 493-512

Kelly G, Prasannan S, Daniell S, Fleming K, Frankel G, Dougan G, Connerton I, Matthews S (1999) Structure of the cell-adhesion fragment of intimin from enteropathogenic *Escherichia coli*. *Nature Structural Biology* **6:** 313-318

Kemp P, Garcia LR, Molineux IJ (2005) Changes in bacteriophage T7 virion structure at the initiation of infection. *Virology* **340:** 307-317

Kholodii GY, Mindlin SZ (1985) Integration of Related Prophages Lambda,Phi-80 and Their Hybrid Lambda-Att80 into Wild-Type Escherichia-Coli Chromosome at Secondary Attachment Sites. *Genetika* **21:** 46-53

Kim YS, Ho SB (2010) Intestinal goblet cells and mucins in health and disease: recent insights and progress. *Current gastroenterology reports* **12:** 319-330

King J, Casjens S (1974) Catalytic Head Assembling Protein in Virus Morphogenesis. *Nature* **251:** 112-119

Klockgether J, Reva O, Larbig K, Tummler B (2004) Sequence analysis of the mobile genome island pKLC102 of *Pseudomonas aeruginosa* C. *Journal of Bacteriology* **186:** 518-534

Kornitzer D, Altuvia S, Oppenheim AB (1991) The Activity of the Ciii Regulator of Lambdoid Bacteriophages Resides within a 24-Amino Acid Protein Domain. *Proceedings of the National Academy of Sciences of the United States of America* **88:** 5217-5221

Kornitzer D, Teff D, Altuvia S, Oppenheim AB (1989) Genetic analysis of bacteriophage lambda cIII gene: mRNA structural requirements for translation initiation. *Journal of Bacteriology* **171:** 2563-2572

Kossykh VG, Schlagman SL, Hattman S (1995) Phage T4 DNA [N6-adenine]methyltransferase. Overexpression, purification, and characterization. *Journal of Biological Chemistry* **270:** 14389-14393

Kropinski AM (2000) Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. *Journal of Bacteriology* **182:** 6066-6074

Kropinski AM, Borodovsky M, Carver TJ, Cerdeno-Tarraga AM, Darling A, Lomsadze A, Mahadevan P, Stothard P, Seto D, Van Domselaar G, Wishart DS (2009) In silico identification of genes in bacteriophage DNA. *Methods in Molecular Biology* **502:** 57-89

Kropinski AM, Krylov V, Pleteneva E, Shaburova O, Bourkaltseva M, Krylov SV, Miroshnikov K. (2010a) *Pseudomonas* Phage Phi297. State Institute for Genetics and Selection of Industrial Microorganisms, 1St Dorozhnii Projezd, 1, Moscow 113545, Russia.

Kropinski AM, Krylov V, Pleteneva E, Shaburova O, Bourkaltseva M, Krylov SV, Miroshnikov K. (2010b) Pseudomonas phage vB_PaeS_PMG1.

Kruger DH, Bickle TA (1983) Bacteriophage Survival - Multiple Mechanisms for Avoiding the Deoxyribonucleic-Acid Restriction Systems of Their Hosts. *Microbiological Reviews* **47:** 345-360

Krumsiek J, Arnold R, Rattei T (2007) Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23:** 1026-1028

Krupovic M, Dutilh BE, Adriaenssens EM, Wittmann J, Vogensen FK, Sullivan MB, Rumnieks J, Prangishvili D, Lavigne R, Kropinski AM, Klumpp J, Gillis A, Enault F, Edwards RA, Duffy S, Clokie MRC, Barylski J, Ackermann HW, Kuhn JH (2016) Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. *Archives of Virology* **161:** 1095-1099

Kubesch P, Lingner M, Grothues D, Wehsling M, Tummler B (1988) Strategies of *Pseudomonas aeruginosa* to Colonize and to Persist in the Cystic-Fibrosis Lung. *Scandinavian Journal of Gastroenterology* **23:** 77-80

Kuti JL, Moss KM, Nicolau DP, Knauft RF (2004) Empiric treatment of multidrug-resistant *Burkholderia cepacia* lung exacerbation in a patient with cystic fibrosis: Application of pharmacodynamic concepts to meropenem therapy. *Pharmacotherapy* **24:** 1641-1645

Kutter E, De Vos D, Gvasalia G, Alavidze Z, Gogokhia L, Kuhl S, Abedon ST (2010) Phage Therapy in Clinical Practice: Treatment of Human Infections. *Current Pharmaceutical Biotechnology* **11:** 69-86

Kwan T, Liu J, DuBow M, Gros P, Pelletier J (2006) Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. *Journal of Bacteriology* **188:** 1184-1187

Lammens EA, Lavigne R. (2013) Complete genome sequence of F116-like bacteriophages. Biosystems, KU Leuven, Kasteelpark Arenberg 21 - Box 2462, Heverlee 3001, Belgium.

Lander GC, Tang L, Casjens SR, Gilcrease EB, Prevelige P, Poliakov A, Potter CS, Carragher B, Johnson JE (2006) The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science* **312:** 1791-1795

Langley R, Kenna DT, Vandamme P, Ure R, Govan JRW (2003) Lysogeny and bacteriophage host range within the *Burkholderia cepacia* complex. *Journal of Medical Microbiology* **52:** 483-490

Latino L, Essoh C, Blouin Y, Thien HV, Pourcel C (2014) A novel *Pseudomonas aeruginosa* Bacteriophage, Ab31, a Chimera Formed from Temperate Phage PAJU2 and P. putida Lytic Phage AF: Characteristics and Mechanism of Bacterial Resistance. *PloS One* **9**

Ledson MJ, Gallagher MJ, Jackson M, Hart CA, Walshaw MJ (2002) Outcome of *Burkholderia cepacia* colonisation in an adult cystic fibrosis centre. *Thorax* **57:** 142-145

Ledson MJ, Gallagher MJ, Walshaw MJ (1998) Chronic *Burkholderia cepacia* bronchiectasis in a non-cystic fibrosis individual. *Thorax* **53:** 430-432

Lee JU, Beveridge TJ (2001) Interaction between iron and *Pseudomonas aeruginosa* biofilms attached to Sepharose surfaces. *Chemical Geology* **180:** 67-80

Lee KK, Gan L, Tsuruta H, Hendrix RW, Duda RL, Johnson JE (2004) Evidence that a local refolding event triggers maturation of HK97 bacteriophage capsid. *Journal of Molecular Biology* **340:** 419-433

Leong D, Murphy JR (1985) Characterization of the diphtheria tox transcript in *Corynebacterium diphtheriae* and *Escherichia coli*. *Journal of Bacteriology* **163:** 1114-1119

LiPuma JJ (2010) The Changing Microbial Epidemiology in Cystic Fibrosis. *Clinical Microbiology Reviews* **23:** 299-+

Liu L, Li YH, Li SL, Hu N, He YM, Pong R, Lin DN, Lu LH, Law M (2012) Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*

Livraghi A, Randell SH (2007) Cystic fibrosis and other respiratory diseases of impaired mucus clearance. *Toxicologic Pathology* **35:** 116-129

Lopes A, Amarir-Bouhram J, Faure G, Petit MA, Guerois R (2010) Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. *Nucleic Acids Research* **38:** 3952-3962

Lukacik P, Barnard TJ, Keller PW, Chaturvedi KS, Seddiki N, Fairman J, Noinaj N, Kirby TL, Henderson JP, Steven AC, Hinnebusch BJ, Buchanan SK (2012) Structural engineering of a phage lysin that targets Gram-negative pathogens. *Proceedings of the National Academy of Sciences of the United States of America* **109:** 9857-9862

Lukashin AV, Borodovsky M (1998) GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Research* **26:** 1107-1115

Lyczak JB, Cannon CL, Pier GB (2002) Lung infections associated with cystic fibrosis. *Clinical Microbiology Reviews* **15:** 194-222

Lynch KH, Stothard P, Dennis JJ (2010) Genomic analysis and relatedness of P2-like phages of the *Burkholderia cepacia* complex. *BMC Genomics* **11**

MacDonald KD, McKenzie KR, Zeitlin PL (2007) Cystic fibrosis transmembrane regulator protein mutations: 'class' opportunity for novel drug innovation. *Paediatric Drugs* **9:** 1-10

Magnuson RD (2007) Hypothetical functions of toxin-antitoxin systems. *Journal of Bacteriology* **189:** 6089-6092

Magoc T, Pabinger S, Canzar S, Liu XY, Su Q, Puiu D, Tallon LJ, Salzberg SL (2013) GAGE-B: an evaluation of genome assemblers for bacterial organisms. *Bioinformatics* **29:** 1718-1725

Mahenthiralingam E, Baldwin A, Dowson CG (2008) *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. *Journal of Applied Microbiology* **104:** 1539-1551

Mahenthiralingam E, Baldwin A, Vandamme P (2002) *Burkholderia cepacia* complex infection in patients with cystic fibrosis. *Journal of Medical Microbiology* **51:** 533-538

Mahenthiralingam E, Urban TA, Goldberg JB (2005) The multifarious, multireplicon *Burkholderia cepacia* complex. *Nature Reviews Microbiology* **3:** 144-156

Mahenthiralingam E, Vandamme P, Campbell ME, Henry DA, Gravelle AM, Wong LTK, Davidson AGF, Wilcox PG, Nakielna B, Speert DP (2001) Infection with *Burkholderia cepacia* complex genomovars in patients with cystic fibrosis: Virulent transmissible strains of genomovar III can replace *Burkholderia multivorans*. *Clinical Infectious Diseases* **33:** 1469-1475

Maida I, Fondi M, Orlandini V, Emiliani G, Papaleo MC, Perrin E, Fani R (2014) Origin, duplication and reshuffling of plasmid genes: Insights from *Burkholderia vietnamiensis* G4 genome. *Genomics* **103:** 229-238

Matsumoto H, Itoh Y, Ohta S, Terawaki Y (1986) A generalized transducing phage of *Pseudomonas cepacia*. *Journal of Genetic Microbiology* **132:** 2583-2586

Matsushiro A, Sato K, Miyamoto H, Yamamura T, Honda T (1999) Induction of prophages of enterohemorrhagic *Escherichia coli* O157 : H7 with norfloxacin. *Journal of Bacteriology* **181:** 2257-2260

Mattern IE, van Winden MP, Rorsch A (1965) The range of action of genes controlling radiation sensitivity in *Escherichia coli*. *Mutation Research* **2:** 111-131

Maya O, Flores V, Guarneros G. (2012) Complete genome sequence of *Pseudomonas aeruginosa* temperate bacteriophage H66. NCBI.

McCallum SJ, Gallagher MJ, Corkill JE, Hart CA, Ledson MJ, Walshaw MJ (2002) Spread of an epidemic *Pseudomonas aeruginosa* strain from a patient with cystic fibrosis (CF) to non-CF relatives. *Thorax* **57:** 559-560

Mcconnell MR, Foster BD, Davis DP, Kat B, Blair JG, Long RA, Steed MM (1986) A Spontaneously Produced Lipopolysaccharide Biosynthetic Defect Which Causes Both Pleiotropic Phage Resistance and Mucoid Colony Morphology in Salmonella-Anatum. *Microbios* **48:** 135-158

McDonald JE, Smith DL, Fogg PCM, McCarthy AJ, Allison HE (2010) High-Throughput Method for Rapid Induction of Prophages from Lysogens and Its Application in the Study of Shiga Toxin-Encoding Escherichia coli Strains. *Applied and Environmental Microbiology* **76:** 2360-2365

McMahon SA, Miller JL, Lawton JA, Kerkow DE, Hodes A, Marti-Renom MA, Doulatov S, Narayanan E, Sali A, Miller JF (2005) The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nature Structural & Molecular Biology* **12:** 886-892

Mello LV, Chen X, Rigden DJ (2010) Mining metagenomic data for novel domains: BACON, a new carbohydrate-binding module. *FEBS Letters* **584:** 2421-2426

Meyer JM (2000) Pyoverdines: pigments, siderophores and potential taxonomic markers of fluorescent *Pseudomonas* species. *Archives of Microbiology* **174:** 135-142

Meynell EW (1961) A phage, phi chi, which attacks motile bacteria. *Journal of Genetic Microbiology* **25:** 253-290

Michalowski CB, Little JW (2005) Positive autoregulation of cI is a dispensable feature of the phage lambda gene regulatory circuitry. *Journal of Bacteriology* **187:** 6430-6442

Miller SCM, LiPuma JJ, Parke JL (2002) Culture-based and non-growth-dependent detection of the *Burkholderia cepacia* complex in soil environments. *Applied and Environmental Microbiology* **68:** 3750-3758

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution* **60:** 174-182

Morgan GJ, Hatfull GF, Casjens S, Hendrix RW (2002) Bacteriophage Mu genome sequence: Analysis and comparison with Mu-like prophages in *Haemophilus, Neisseria and Deinococcus*. *Journal of Molecular Biology* **317:** 337-359

Morillas HN, Zariwala M, Knowles MR (2007) Genetic causes of bronchiectasis: Primary ciliary dyskinesia. *Respiration* **74:** 252-263

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92:** 255-264

Murphy KC (2012) Phage Recombinases and Their Applications. *Advances in Virus Research, Vol 83: Bacteriophages, Pt B* **83:** 367-414

Naderer M, Brust JR, Knowle D, Blumenthal RM (2002) Mobility of a restriction-modification system revealed by its genetic contexts in three hosts. *Journal of Bacteriology* **184:** 2411-2419

Nakayama K, Kanaya S, Ohnishi M, Terawaki Y, Hayashi T (1999) The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages. *Molecular Microbiology* **31:** 399-419

Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29:** 2933-2935

NCBI. (2016) NCBI genome. NCBI.

Nelson DC, Schmelcher M, Rodriguez-Rubio L, Klumpp J, Pritchard DG, Dong SL, Donovan DM (2012) Endolysins as Antimicrobials. *Advances in Virus Research, Vol 83: Bacteriophages, Pt B* **83:** 299-365

NHS.uk. (2015) Bronchiectasis. Vol. 2016.

NHS.uk. (2016) Cystic fibrosis - Treatment. Vol. 2016.

NIH. (2016) What Is COPD? , Vol. 2016.

Nimmich W (1994) Detection of *Escherichia coli* K95 Strains by Bacteriophages. *Journal of Clinical Microbiology* **32:** 2843-2845

Nishiyama E, Ohtsubo Y, Nagata Y, Tsuda M (2010) Identification of *Burkholderia multivorans* ATCC 17616 genes induced in soil environment by in vivo expression technology. *Environmental Microbiology* **12:** 2539-2558

Nnalue NA, Newton S, Stocker BAD (1990) Lysogenization of Salmonella-Choleraesuis by Phage-14 Increases Average Length of O-Antigen Chains, Serum Resistance and Intraperitoneal Mouse Virulence. *Microbial Pathogenesis* **8:** 393-402

Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, Kambal A, Monaco CL, Zhao G, Fleshner P, Stappenbeck TS, McGovern DPB, Keshavarzian A, Mutlu EA, Sauk J, Gevers D, Xavier RJ, Wang D, Parkes M, Virgin HW (2015) Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* **160:** 447-460

Nzula S, Vandamme P, Govan JRW (2000) Sensitivity of the *Burkholderia cepacia* complex and *Pseudomonas aeruginosa* to transducing bacteriophages. *FEMS Immunology and Medical Microbiology* **28:** 307-312

Nzula S, Vandamme P, Govan JRW (2002) Influence of taxonomic status on the in vitro antimicrobial susceptibility of the *Burkholderia cepacia* complex. *Journal of Antimicrobial Chemotherapy* **50:** 265-269

O'Sullivan BP, Freedman SD (2009) Cystic fibrosis. *Lancet* **373:** 1891-1904

O'Sullivan LA, Weightman AJ, Jones TH, Marchbank AM, Tiedje JM, Mahenthiralingam E (2007) Identifying the genetic basis of ecologically and biotechnologically useful functions of the bacterium *Burkholderia vietnamiensis*. *Environmental Microbiology* **9:** 1017-1034

Ogura T, Hiraga S (1983) Mini-F Plasmid Genes That Couple Host-Cell Division to Plasmid Proliferation. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **80:** 4784-4788

Ohgaki N (1994) Bacterial biofilm in chronic airway infection. *Kansenshogaku Zasshi Journal of the Japanese Association for Infectious Diseases* **68:** 138-151

Ojeniyi B, Birchandersen A, Mansa B, Rosdahl VT, Hoiby N (1991) Morphology of *Pseudomonas aeruginosa* Phages from the Sputum of Cystic-Fibrosis Patients and from the Phage Typing Set - an Electron-Microscopy Study. *APMIS* **99:** 925-930

Olivieri D, Ciaccia A, Marangio E, Marsico S, Todisco T, Delvita M (1991) Role of Bromhexine in Exacerbations of Bronchiectasis - Double-Blind Randomized Multicenter Study Versus Placebo. *Respiration* **58:** 117-121

Onofre-Lemus J, Hernandez-Lucas I, Girard L, Caballero-Mellado J (2009) ACC (1-Aminocyclopropane-1-Carboxylate) Deaminase Activity, a Widespread Trait in *Burkholderia* Species, and Its Growth-Promoting Effect on Tomato Plants. *Applied and Environmental Microbiology* **75:** 6581-6590

Orens JB, Estenne M, Arcasoy S, Conte JV, Corris P, Egan JJ, Egan T, Keshavjee S, Knoop C, Kotloff R, Martinez FJ, Nathan S, Palmer S, Patterson A, Singer L, Snell G, Studer S, Vachiery JL, Glanville AR (2006) International Guidelines for the Selection of Lung Transplant candidates: 2006 update - A consensus report from the pulmonary scientific council of the International Society for Heart and Lung Transplantation. *Journal of Heart and Lung Transplantation* **25:** 745-755

Pacheco SV, Gonzalez OG, Contreras GLP (1997) The lom gene of bacteriophage lambda is involved in *Escherichia coli* K12 adhesion to human buccal epithelial cells. *FEMS Microbiology Letters* **156:** 129-132

Pankov R, Yamada KM (2002) Fibronectin at a glance. *Journal of Cell Science* **115:** 3861-3863

Pasteur MC, Helliwell SM, Houghton S, Webb SC, Foweraker JE, Coulden RA, Flower CD, Bilton D, Keogan MT (2000) An investigation into causative factors in patients with bronchiectasis. *American Journal of Respiratory and Critical Care Medicine* **162:** 1277-1284

Payne GW, Vandamme P, Morgan SH, LiPuma JJ, Coenye T, Weightman AJ, Jones TH, Mahenthiralingam E (2005) Development of a *recA* gene-based identification approach for the the entire *Burkholderia* genus. *Applied and Environmental Microbiology* **71:** 3917-3927

Pecota DC, Wood TK (1996) Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. *Journal of Bacteriology* **178:** 2044-2050

Pelkonen S, Aalto J, Finne J (1992) Differential Activities of Bacteriophage Depolymerase on Bacterial Polysaccharide - Binding Is Essential but Degradation Is Inhibitory in Phage Infection of K1-Defective *Escherichia coli*. *Journal of Bacteriology* **174:** 7757-7761

Pell LG, Gasmi-Seabrook G, Morais M, Neudecker P, Kanelis V, Bona D, Donaldson LW, Edwards AM, Howell PL, Davidson AR (2010) The solution structure of the C-terminal Ig-like domain of the bacteriophage λ tail tube protein. *Journal of Molecular Biology* **403:** 468-479

Pemberton JM (1973) F116: a DNA bacteriophage specific for the pili of *Pseudomonas aeruginosa* strain PAO. *Virology* **55:** 558-560

Peng Y, Leung HC, Yiu SM, Chin FY (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28:** 1420-1428

Pevzner PA, Tang HX, Waterman MS (2001) An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 9748-9753

Plunkett G, Rose DJ, Durfee TJ, Blattner FR (1999) Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157 : H7: Shiga toxin as a phage late-gene product. *Journal of Bacteriology* **181:** 1767-1778

Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jacobs WR, Hendrix RW, Lawrence JG, Hatfull GF, Hunters SEAP, Hunters SEAP, Course MG (2015) Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *Elife* **4**

Pragman AA, Kim HB, Reilly CS, Wendt C, Isaacson RE (2012) The Lung Microbiome in Moderate and Severe Chronic Obstructive Pulmonary Disease. *PloS One* **7**

Proft T, Moffatt SL, Berkahn CJ, Fraser JD (1999) Identification and characterization of novel superantigens from *Streptococcus pyogenes*. *Journal of Experimental Medicine* **189:** 89-101

Ptashne M (2004) *A genetic switch : phage lambda revisited*, 3rd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Puck TT, Garen A, Cline J (1951) The mechanism of virus attachment to host cells. I. The role of ions in the primary reaction. *The Journal of Experimental Medicine* **93:** 65-88

Purcell P, Jary H, Perry A, Perry JD, Stewart CJ, Nelson A, Lanyon C, Smith DL, Cummings SP, De Soyza A (2014a) Polymicrobial airway bacterial communities in adult bronchiectasis patients. *Bmc Microbiology* **14**

Purcell P, Jary H, Perry A, Perry JD, Stewart CJ, Nelson A, Lanyon C, Smith DL, Cummings SP, De Soyza A (2014b) Polymicrobial airway bacterial communities in adult bronchiectasis patients. *BMC Microbiology* **14:** 130

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**

Rakhuba DV, Kolomiets EI, Dey ES, Novik GI (2010) Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Polish Journal of Microbiology* **59:** 145-155

Ramalho AS, Beck S, Meyer M, Penque D, Cutting GR, Amaral MD (2002) Five percent of normal cystic fibrosis transmembrane conductance regulator mRNA ameliorates the severity of pulmonary disease in cystic fibrosis. *American Journal of Respiratory Cell and Molecular Biology* **27:** 619-627

Ramos FL, Krahnke JS, Kim V (2014) Clinical issues of mucus accumulation in COPD. *International Journal of Chronic Obstructive Pulmonary Disease* **9:** 139-150

Ramsey BW (1996) Management of pulmonary disease in patients with cystic fibrosis *New England Journal of Medicine* **335:** 1167-1167

Raoult D, Forterre P (2008) Redefining viruses: lessons from Mimivirus. *Nature Reviews Microbiology* **6:** 315-319

Ray K, Ma J, Oram M, Lakowicz JR, Black LW (2010) Single-molecule and FRET fluorescence correlation spectroscopy analyses of phage DNA packaging: colocalization of packaged phage T4 DNA ends within the capsid. *Journal of Molecular Biology* **395:** 1102-1113

Reese J, Urry, L., Cain, M., Wasserman, S., Minorsky, P., and Jackson, R. (2011) *Campbell Biology*, 9th edition edn. San Francisco, USA: Pearson Benjamin Cummings.

Reichardt W, Muller-Alouf H, Kohler W (1993) Erythrogenic toxin type A (ETA): epidemiological analysis of gene distribution and protein formation in clinical *Streptococcus pyogenes* strains causing scarlet fever and the streptococcal toxic shock-like syndrome (TSLS). *Zentralblatt fur Bakteriologie : International Journal of Medical Microbiology* **279:** 283-293

Reis VM, Estrada-de los Santos P, Tenorio-Salgado S, Vogel J, Stoffels M, Guyon S, Mavingui P, Baldani VLD, Schmid M, Baldani JI, Balandreau J, Hartmann A, Caballero-Mellado J (2004) *Burkholderia tropica* sp nov., a novel nitrogen-fixing, plant-associated bacterium. *International Journal of Systematic and Evolutionary Microbiology* **54:** 2155-2162

Ren CL, Konstan MW, Yegin A, Rasouliyan L, Trzaskoma B, Morgan WJ, Regelmann W, Grp SA, Study E (2012) Multiple antibiotic-resistant *Pseudomonas aeruginosa* and lung function decline in patients with cystic fibrosis. *Journal of Cystic Fibrosis* **11:** 293-299

Reyes EAP, Bale MJ, Cannon WH, Matsen JM (1981) Identification of *Pseudomonas aeruginosa* by Pyocyanin Production on Tech Agar. *Journal of Clinical Microbiology* **13:** 456-458

Riede I, Drexler K, Eschbach ML (1985) The nucleotide sequences of the tail fiber gene 36 of bacteriophage T2 and of genes 36 of the T-even type *Escherichia coli* phages K3 and Ox2. *Nucleic Acids Research* **13:** 605-616

Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, et al. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245:** 1066-1073

Roberts DE, Dhillon DP, Porteous E, Cole PJ (1984) Treatment of *Pseudomonas* Infestation in Bronchiectasis with the Quinolone Ciprofloxacin. *Thorax* **39:** 719-720

Rodley PD, Romling U, Tummler B (1995) A Physical Genome Map of the *Burkholderia cepacia* Type Strain. *Molecular Microbiology* **17:** 57-67

Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Bruce KD (2004) Characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16S ribosomal DNA terminal restriction fragment length polymorphism profiling. *Journal of Clinical Microbiology* **42:** 5176-5183

Rogers GB, Hart CA, Mason JR, Hughes M, Walshaw MJ, Bruce KD (2003) Bacterial diversity in cases of lung infection in cystic fibrosis patients: 16S ribosomal DNA (rDNA) length heterogeneity PCR and 16S rDNA terminal restriction fragment length polymorphism profiling. *Journal of Clinical Microbiology* **41:** 3548-3558

Rogers GB, Hoffman LR, Whiteley M, Daniels TWV, Carroll MP, Bruce KD (2010) Revealing the dynamics of polymicrobial infections: implications for antibiotic therapy. *Trends in Microbiology* **18:** 357-364

Rogers GB, van der Gast CJ, Cuthbertson L, Thomson SK, Bruce KD, Martin ML, Serisier DJ (2013) Clinical measures of disease in adult non-CF bronchiectasis correlate with airway microbiota composition. *Thorax* **68:** 731-737

Rolain JM, Francois P, Hernandez D, Bittar F, Richet H, Fournous G, Mattenberger Y, Bosdure E, Stremler N, Dubus JC, Sarles J, Reynaud-Gaubert M, Boniface S, Schrenzel J, Raoult D (2009) Genomic analysis of an emerging multiresistant *Staphylococcus aureus* strain rapidly spreading in cystic fibrosis patients revealed the presence of an antibiotic inducible bacteriophage. *Biology Direct* **4**

Roncero C, Darzins A, Casadaban MJ (1990) *Pseudomonas aeruginosa* transposable bacteriophages D3112 and B3 require pili and surface growth for adsorption. *Journal of Bacteriology* **172:** 1899-1904

Ronning CM, Losada L, Brinkac L, Inman J, Ulrich RL, Schell M, Nierman WC, DeShazer D (2010) Genetic and phenotypic diversity in *Burkholderia*: contributions by prophage and phage-like elements. *BMC Microbiology* **10**

Rose MC, Brown CF, Jacoby JZ, 3rd, Lynn WS, Kaufman B (1987) Biochemical properties of tracheobronchial mucins from cystic fibrosis and non-cystic fibrosis individuals. *Pediatric Research* **22:** 545-551

Rose MC, Nickola TJ, Voynow JA (2001) Airway mucus obstruction: mucin glycoproteins, MUC gene regulation and goblet cell hyperplasia. *American Journal of Respiratory Cell and Molecular Biology* **25:** 533-537

Rosenstein BJ, Cutting GR, Panel CFFC (1998) The diagnosis of cystic fibrosis: A consensus statement. *Journal of Pediatrics* **132:** 589-595

Rudkjobing VB, Thomsen TR, Alhede M, Kragh KN, Nielsen PH, Johansen UR, Givskov M, Hoiby N, Bjarnsholt T (2011) True Microbiota Involved in Chronic Lung Infection of Cystic Fibrosis Patients Found by Culturing and 16S rRNA Gene Analysis. *Journal of Clinical Microbiology* **49:** 4352-4355

Sako T, Tsuchida N (1983) Nucleotide sequence of the *staphylokinase* gene from *Staphylococcus aureus*. *Nucleic Acids Research* **11:** 7679-7693

Scandella D, Arber W (1974) An *Escherichia coli* mutant which inhibits the injection of phage lambda DNA. *Virology* **58:** 504-513

Schade SZ, Adler J, Ris H (1967) How bacteriophage chi attacks motile bacteria. *Journal of Virology* **1:** 599-609

Schwartz M (1976) Adsorption of Coliphage Lambda to Its Host - Effect of Variations in Surface Density of Receptor and in Phage-Receptor Affinity. *Journal of Molecular Biology* **103:** 521-536

Scott D, Ely B (2015) Comparison of Genome Sequencing Technology and Assembly Methods for the Analysis of a GC-Rich Bacterial Genome. *Current Microbiology* **70:** 338-344

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30:** 2068-2069

Sepulveda-Robles O, Kameyama L, Guarneros G (2012) High Diversity and Novel Species of *Pseudomonas aeruginosa* Bacteriophages. *Applied and Environmental Microbiology* **78:** 4510-4515

Sestini P, Renzoni E, Robinson S, Poole P, Ram FS (2002) Short-acting beta 2 agonists for stable chronic obstructive pulmonary disease. *Cochrane Database Syst Rev***:** CD001495

Sethi S, Cote C (2011) Bronchodilator combination therapy for the treatment of chronic obstructive pulmonary disease. *Current Clinical Pharmacology* **6:** 48-61

Setlow JK, Boling ME, Allison DP, Beattie KL (1973) Relationship between Prophage Induction and Transformation in *Haemophilus influenzae*. *Journal of Bacteriology* **115:** 153-161

Shafiq I, Carroll MP, Nightingale JA, Daniels TV (2011) Cepacia syndrome in a cystic fibrosis patient colonised with *Burkholderia multivorans*. *BMJ Case Reports* **2011**

Shao YP, Wang IN (2008) Bacteriophage adsorption rate and optimal lysis time. *Genetics* **180:** 471-482

Sharan SK, Thomason LC, Kuznetsov SG, Court DL (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nature Protocols* **4:** 206-223

Sheppard DN, Welsh MJ (1999) Structure and function of the CFTR chloride channel. *Physiological Reviews* **79:** S23-45

Sibley CD, Rabin H, Surette MG (2006) Cystic fibrosis: a polymicrobial infectious disease. *Future Microbiology* **1:** 53-61

Silveira CB, Rohwer FL (2016) Piggyback-the-Winner in host-associated microbial communities. *npj Biofilms and Microbiomes* **2:** 16010

Skorupski K, Tomaschewski J, Ruger W, Simon LD (1988) A Bacteriophage-T4 Gene Which Functions to Inhibit *Escherichia coli* Lon Protease. *Journal of Bacteriology* **170:** 3016-3024

Skurray RA, Hancock REW, Reeves P (1974) Con-Mutants - Class of Mutants in *Escherichia coli* K-12 Lacking a Major Cell wall Protein and Defective in Conjugation and Adsorption of a Bacteriophage. *Journal of Bacteriology* **119:** 726-735

Slideplayer.inc. (2016) Next generation sequencing: Illuminia/Solexa sequencing. Slideplayer.com, Vol. 2016.

Smith DL, Rooks DJ, Fogg PCM, Darby AC, Thomson NR, McCarthy AJ, Allison HE (2012) Comparative genomics of Shiga toxin encoding bacteriophages. *BMC Genomics* **13**

Smits SL, Bodewes R, Ruiz-Gonzalez A, Baumgartner W, Koopmans MP, Osterhaus ADME, Schurch AC (2014) Assembly of viral genomes from metagenomes. *Frontiers in Microbiology* **5**

Smits SL, Bodewes R, Ruiz-Gonzalez A, Baumgartner W, Koopmans MP, Osterhaus ADME, Schurch AC (2015) Recovering full-length viral genomes from metagenomes. *Frontiers in Microbiology* **6**

Snyder L (1995) Phage-Exclusion Enzymes - a Bonanza of Biochemical and Cell Biology Reagents. *Molecular Microbiology* **15:** 415-420

Sorde R, Pahissa A, Rello J (2011) Management of refractory *Pseudomonas aeruginosa* infection in cystic fibrosis. *Infection and Drug Resistance* **4:** 31-41

Spakowitz AJ, Wang ZG (2005) DNA packaging in bacteriophage: Is twist important? *Biophysical Journal* **88:** 3912-3923

Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao QX, Koonin EV, Davis RW (1998) Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. *Science* **282:** 754-759

Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FSL, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GKS, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock REW, Lory S, Olson MV (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406:** 959-964

Summer EJ, Gill JJ, Upton C, Gonzalez CF, Young R (2007) Role of phages in the pathogenesis of *Burkholderia*, or 'Where are the toxin genes in *Burkholderia* phages?'. *Current Opinion in Microbiology* **10:** 410-417

Summer EJ, Gonzalez CF, Carlisle T, Mebane LM, Cass AM, Savva CG, LiPuma JJ, Young R (2004) *Burkholderia cenocepacia* phage BcepMu and a family of Mu-like phages encoding potential pathogenesis factors. *Journal of Molecular Biology* **340:** 49-65

Sun CL, Barrangou R, Thomas BC, Horvath P, Fremaux C, Banfield JF (2013) Phage mutations in response to CRISPR diversification in a bacterial population. *Environmental Microbiology* **15:** 463-470

Suttle CA (2007) Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology* **5:** 801-812

Tabacchioni S, Ferri L, Manno G, Mentasti M, Cocchi P, Campana S, Ravenni N, Taccetti G, Dalmastri C, Chiarini L, Bevivino A, Fani R (2008) Use of the gyrB gene to discriminate among species of the *Burkholderia cepacia* complex. *FEMS Microbiology Letters* **281:** 175-182

Talbi C, Argandona M, Salvador M, Alche JD, Vargas C, Bedmar EJ, Delgado MJ (2013) *Burkholderia phymatum* improves salt tolerance of symbiotic nitrogen fixation in Phaseolus vulgaris. *Plant and Soil* **367:** 673-685

Tariq MA, Everest FLC, Cowley LA, De Soyza A, Holt GS, Bridge SH, Perry A, Perry JD, Bourke SJ, Cummings SP, Lanyon CV, Barr JJ, Smith DL (2015) A metagenomic approach to characterize temperate bacteriophage populations from Cystic Fibrosis and non-Cystic Fibrosis bronchiectasis patients. *Frontiers in Microbiology* **6**

Tayeb LA, Lefevre M, Passet V, Diancourt L, Brisse S, Grimont PAD (2008) Comparative phylogenies of *Burkholderia, Ralstonia, Comamonas, Brevundimonas* and related organisms derived from *rpoB, gyrB* and *rrs* gene sequences. *Research in Microbiology* **159:** 169-177

ten Hacken NHT, Wijkstra PJ, Kerstjens HAM (2007) Treatment of bronchiectasis in adults. *British Medical Journal* **335:** 1089-1093

Terns MP, Terns RM (2011) CRISPR-based adaptive immune systems. *Current Opinion in Microbiology* **14:** 321-327

Thingstad TF, Lignell R (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology* **13:** 19-27

Thisted T, Gerdes K (1992) Mechanism of Post-Segregational Killing by the Hok Sok System of Plasmid R1 - Sok Antisense Rna Regulates Hok Gene-Expression Indirectly through the Overlapping Mok Gene. *Journal of Molecular Biology* **223:** 41-54

Thomas CM, Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology* **3:** 711-721

Thornton DJ, Rousseau K, McGuckin MA (2008) Structure and function of the polymeric mucins in airways mucus. *Annual Review of Physiology* **70:** 459-486

Thorpe HM, Smith MCM (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 5505-5510

Trepanier S, Prince A, Huletsky A (1997) Characterization of the penA and penR genes of *Burkholderia cepacia* 249 which encode the chromosomal class a penicillinase and its LysR-type transcriptional regulator. *Antimicrobial Agents and Chemotherapy* **41:** 2399-2405

Turton JF, Kaufmann ME, Mustafa N, Kawa S, Clode FE, Pitt TL (2003) Molecular comparison of isolates of *Burkholderia multivorans* from patients with cystic fibrosis in the United Kingdom. *Journal of Clinical Microbiology* **41:** 5750-5754

Twort FW (1915) An Investigation on the nature of ultra-microscopic viruses. *Lancet* **186:** 1241-1243

Valenick LV, Hsia HC, Schwarzbauer JE (2005) Fibronectin fragmentation promotes alpha 4 beta 1 integrin-mediated contraction of a fibrin-fibronectin provisional matrix. *Experimental Cell Research* **309:** 48-55

Valenza G, Tappe D, Turnwald D, Frosch M, Konig C, Hebestreit H, Abele-Horn M (2008) Prevalence and antimicrobial susceptibility of microorganisms isolated from sputa of patients with cystic fibrosis. *Journal of Cystic Fibrosis* **7:** 123-127

Valpuesta JM, Carrascosa JL (1994) Structure of Viral Connectors and Their Function in Bacteriophage Assembly and DNA Packaging. *Quarterly Reviews of Biophysics* **27:** 107-155

Van Valen L (1973) A new evolutionary law. *Evolutionary Theory* **1:** 1-30

Viallard V, Poirier I, Cournoyer B, Haurat J, Wiebkin S, Ophel-Keller K, Balandreau J (1998) *Burkholderia graminis* sp, nov., a rhizospheric *Burkholderia* species, and reassessment of *[Pseudomonas] phenazinium, [Pseudomonas] pyrrocinia* and *[Pseudomonas] glathei* as *Burkholderia*. *International Journal of Systematic Bacteriology* **48:** 549-563

Villion M, Moineau S (2013) VIROLOGY Phages hijack a host's defence. *Nature* **494:** 433-434

Vogel H, Jahnig F (1986) Models for the structure of outer-membrane proteins of *Escherichia coli* derived from raman spectroscopy and prediction methods. *Journal of Molecular Biology* **190:** 191-199

Voynow JA, Rubin BK (2009) Mucins, Mucus, and Sputum. *Chest* **135:** 505-512

Waldor MK, Mekalanos JJ (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272:** 1910-1914

Walters MC, 3rd, Roe F, Bugnicourt A, Franklin MJ, Stewart PS (2003) Contributions of antibiotic penetration, oxygen limitation, and low metabolic activity to tolerance of *Pseudomonas aeruginosa* biofilms to ciprofloxacin and tobramycin. *Antimicrobial Agents and Chemotherapy* **47:** 317-323

Wang PW, Chu L, Guttman DS (2004) Complete sequence and evolutionary genomic analysis of the *Pseudomonas aeruginosa* transposable bacteriophage D3112. *Journal of Bacteriology* **186:** 400-410

Watanabe R, Matsumoto T, Sano G, Ishii Y, Tateda K, Sumiyama Y, Uchiyama J, Sakurai S, Matsuzaki S, Imai S, Yamaguchi K (2007) Efficacy of bacteriophage therapy against gut-derived sepsis caused by *Pseudomonas aeruginosa* in mice. *Antimicrobial Agents and Chemotherapy* **51:** 446-452

Weaver VB, Kolter R (2004) *Burkholderia* spp. alter *Pseudomonas aeruginosa* physiology through iron sequestration. *Journal of Bacteriology* **186:** 2376-2384

Weitz JS, Poisot T, Meyer JR, Flores CO, Valverde S, Sullivan MB, Hochberg ME (2013) Phage-bacteria infection networks. *Trends in Microbiology* **21:** 82-91

Welsh MJ, Smith AE (1993) Molecular Mechanisms of Cftr Chloride Channel Dysfunction in Cystic-Fibrosis. *Cell* **73:** 1251-1254

Whitchurch CB, Tolker-Nielsen T, Ragas PC, Mattick JS (2002) Extracellular DNA required for bacterial biofilm formation. *Science* **295:** 1487

Williams HT (2013) Phage-induced diversification improves host evolvability. *BMC Evolutionary Biology* **13:** 17

Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, Tammadoni S, Nosrat B, Conrad D, Rohwer F (2009) Metagenomic Analysis of Respiratory Tract DNA Viral Communities in Cystic Fibrosis and Non-Cystic Fibrosis Individuals. *PloS One* **4**

Wilson R, Pitt T, Taylor G, Watson D, MacDermot J, Sykes D, Roberts D, Cole P (1987) Pyocyanin and 1-hydroxyphenazine produced by *Pseudomonas aeruginosa* inhibit the beating of human respiratory cilia in vitro. *Journal of Clinical Investigation* **79:** 221-229

Winstanley C, Langille MGI, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, Sanschagrin F, Thomson NR, Winsor GL, Quail MA, Lennard N, Bignell A, Clarke L, Seeger K, Saunders D, Harris D, Parkhill J, Hancock REW, Brinkman FSL, Levesque RC (2009) Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. *Genome Research* **19:** 12-23

Winstanley C, O'Brien S, Brockhurst MA (2016) *Pseudomonas aeruginosa* Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. *Trends in Microbiology* **24:** 327-337

Xu J, Hendrix RW, Duda RL (2004) Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Molecular Cell* **16:** 11-21

Yabuuchi E, Kosako Y, Oyaizu H, Yano I, Hotta H, Hashimoto Y, Ezaki T, Arakawa M (1992) Proposal of *Burkholderia* Gen-Nov and Transfer of 7 Species of the Genus *Pseudomonas* Homology Group-Ii to the New Genus, with the Type Species *Burkholderia-Cepacia* (Palleroni and Holmes 1981) Comb-Nov. *Microbiology and Immunology* **36:** 1251-1275

Yabuuchi E, Kosako Y, Yano I, Hotta H, Nishiuchi Y (1995) Transfer of 2 *Burkholderia* and an Alcaligenes Species to *Ralstonia* Gen-Nov - Proposal of *Ralstonia-Pickettii* (Ralston, Palleroni and Doudoroff 1973) Comb-Nov, *Ralstonia-Solanacearum* (Smith 1896) Comb-Nov and *Ralstonia-Eutropha* (Davis 1969) Comb-Nov. *Microbiology and Immunology* **39:** 897-904

Yim CF, Lim KS, Low TC (2002) Severe pulmonary hypertension in a patient with bronchiectasis complicated by cor pulmonale and a right-to-left shunt presenting for surgery. *Anaesthesia and Intensive Care* **30:** 467-471

Young R, Wang IN, Roof WD (2000) Phages will out: strategies of host cell lysis. *Trends in Microbiology* **8:** 120-128

Yu DG, Sawitzke JA, Ellis H, Court DL (2003) Recombineering with overlapping single-stranded DNA oligonucleotides: Testing a recombination intermediate. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 7207-7212

Yu F, Mizushima S (1982) Roles of lipopolysaccharide and outer membrane protein OmpC of *Escherichia coli* K-12 in the receptor function for bacteriophage T4. *Journal of Bacteriology* **151:** 718-722

Zahariadis G, Levy MH, Burns JL (2003) Cepacia-like syndrome caused by *Burkholderia multivorans*. *Canadian Journal of Infectious Diseases* **14:** 123-125

Zhang J, Chiodini R, Badr A, Zhang GF (2011) The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics* **38:** 95-109

Zielenski J, Tsui LC (1995) Cystic fibrosis: Genotypic and phenotypic variations. *Annual Review of Genetics* **29:** 777-807

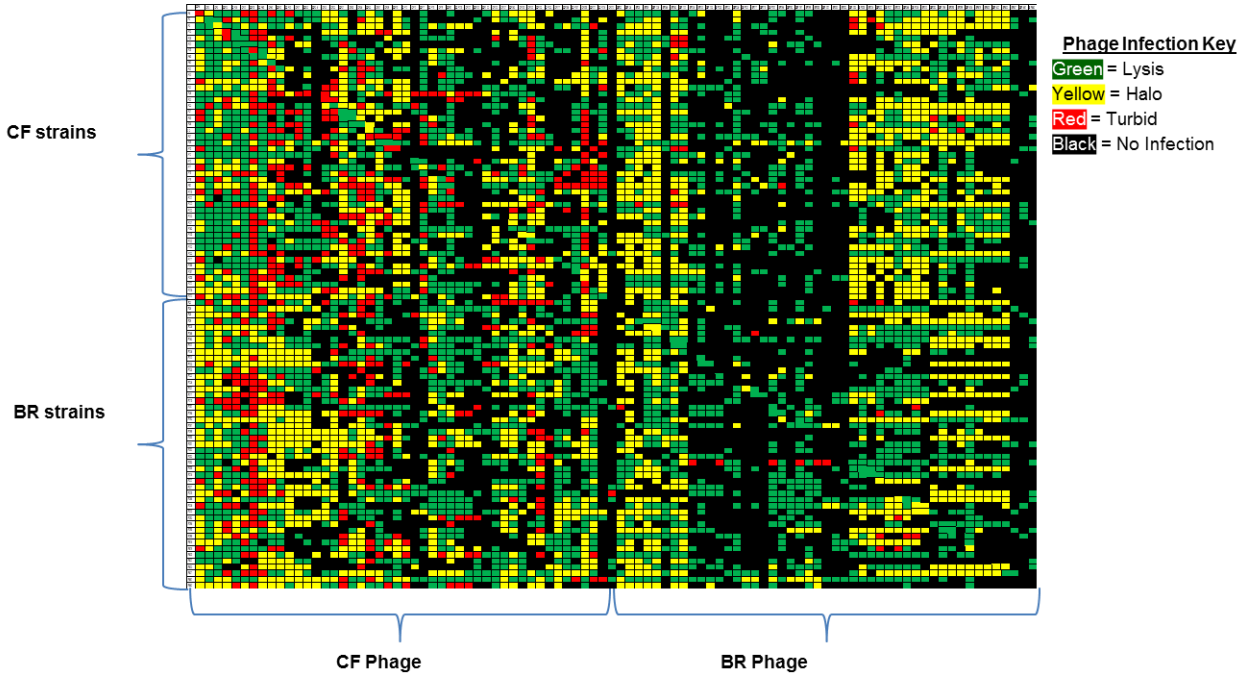
Zissler J (1967) Integration-negative (int) mutants of phage lambda. *Virology* **31:** 189

Zlosnik JEA, Costa PS, Brant R, Mori PYB, Hird TJ, Fraenkel MC, Wilcox PG, Davidson AGF, Speert DP (2011) Mucoid and Nonmucoid *Burkholderia cepacia* Complex Bacteria in Cystic Fibrosis Infections. *American Journal of Respiratory and Critical Care Medicine* **183:** 67-72

Zlosnik JEA, Hird TJ, Fraenkel MC, Moreira LM, Henry DA, Speert DP (2008) Differential mucoid exopolysaccharide production by members of the *Burkholderia cepacia* complex. *Journal of Clinical Microbiology* **46:** 1470-1473

9. Appendices

Appendix 1 Cross infection of *Pseudomonas aeruginosa*



Appendix 1: Cross Infection Profile of CF and BR phages against host range 94 Phage lysates were spotted onto 47 BR and 47 CF *P. aeruginosa* to look at host range of infectivity based on whether the phage lysate was from a CF or BR isolate. Each cell represents a cross infection where the key can be seen denoting observations of infection although titre of lysates is not known. The figure illustrates the infection patterns of cross infection. The results are colour coded for improved visual interpretation.

.

Appendix 2 Table of *Pseudomonas aeruginosa* assemblies

SPAdes 3.1.1				
Sample number and peaks	**N50**	**Number of contigs >=500bp**	**largest contig (bp)**	**Total length >= 0 bp**
1peak1	3759	10	6123	23458
1peak2	20663	3	20663	40708
1	8330	14	25454	57861
2	2064	43	3807	61165
3peak1	4771	7	5928	23329
3peak1+3	6112	9	11150	44310
3peak2	17961	9	17961	29726
3peak2+3	40541	1	40541	41704
3peak3	3903	10	5927	27647
3	8039	11	27661	58313
4	5714	30	16984	80314
5	52800	2	52800	90569
6peak1	13281	34	24866	106624
6peak2	833	3	970	4945
6	61649	3	61649	102039
7peak1	37361	9	37361	59410
7peak1+2	8393	23	37361	84929
7peak1+3	11199	20	37361	85159
7peak2	1915	19	8393	30232
7peak2+3	24302	2	24302	40245
7peak3	2916	15	4301	28142
7	14035	14	37361	95674
8peak1	3112	10	5939	22762
8peak2	6719	10	6719	34082
8peak3/4	11162	7	11162	18743
8	1909	49	5871	130085
9peak1	4493	8	8956	24858

9peak1+2	6292	17	8962	45691
9peak1+3	11383	8	16531	45362
9peak2	7574	10	8752	27450
9peak2+3	40424	3	40424	42309
9peak3	4391	13	6110	30119
9	8363	15	24014	57788
10peak1	7779	15	20434	52312
10peak2	658	2	658	1201
10	9016	14	20434	53448
11	1606	8	2802	10912
12	3603	29	7027	81465
13peak1	4584	6	8739	20223
13peak1+2	3737	16	8479	42625
13peak1+3	31474	7	31474	45404
13peak2	16374	6	16374	23905
13peak2+3	13697	5	14958	43902
13peak3	4904	16	11181	32459
13	8330	17	20288	59285
14peak1	28082	23	38360	165968
14peak1+2	14421	52	38360	262649
14peak1+3	12748	49	38360	223058
14peak2	5379	37	15972	101864
14peak2+3	15924	23	40853	131154
14peak3	4262	22	6802	55673
14	17035	41	38360	292551
15peak1	2500	8	9161	29596
15peak2	8469	17	12344	40155
15peak3	2607	22	6244	38266
15peak4	27956	2	27956	37501
15	62077	6	62077	115756
16peak1	4602	7	10220	21840
16peak1+2	3317	16	8824	41144

16peak1+3	35727	8	35727	47133
16peak2	4937	6	8835	22393
16peak2+3	25451	4	25451	43110
16peak3	2594	13	11219	34183
16	1670	46	7099	138313
17peak1	52664	15	52664	66757
17peak2	7608	9	12623	36976
17	52664	2	52664	90767
18peak1	6843	7	8683	22259
18peak1+2	3519	17	11077	41929
18peak1+3	35225	6	35225	45998
18peak2	5139	8	9977	24970
18peak2+3	2446	20	5045	71323
18peak3	2657	13	10324	32696
18	1366	63	4913	186631
19	23619	4	29993	68794
20peak1	8924	9	11519	31651
20peak2	52662	2	52662	90291
20	61562	6	61562	116309
21peak1	10588	3	10588	22867
21peak1+2	2576	10	8581	46165
21peak1+3	21485	3	21485	42613
21peak2	16810	3	16810	28186
21peak2+3	9753	6	18897	46206
21peak3	5086	8	5086	28332
21	3055	20	11703	62995
22peak1	61772	4	61772	73514
22peak2	566	1	566	1723
22	61772	7	61772	73579
23	4681	21	8473	45998
24peak1	794	2	1977	7649
24peak2	823	1	823	2056

24	4019	6	4019	8815
25	10598	28	44244	157980
26peak1	41525	28	41525	85813
26peak2	7032	7	12605	35917
26	37617	17	41525	109783
27	39705	1	39705	39833
28peak1	4578	7	5925	22235
28peak1+2	2941	19	8356	45017
28peak1+3	41155	4	41155	44910
28peak2	5179	10	10116	30771
28peak2+3	2848	19	4243	62997
28peak3	7807	9	11195	30156
28	3772	24	7654	68669
29peak1	10321	46	27683	123963
29peak1+2	10321	33	27683	145062
29peak1+3	26022	52	33114	213173
29peak2	2540	20	4372	37931
29peak2+3	22903	21	41451	124515
29peak3	52823	2	52823	90459
29	10321	50	28830	233002
30peak1	4358	59	16094	146734
30peak1+2	5801	59	31273	213164
30peak1+3	5570	78	18321	236568
30peak2	4996	29	10935	73859
30peak2+3	26254	25	52758	152514
30peak3	10520	13	36058	88424
30	11205	58	33124	304018
31peak1	38379	36	55177	184505
31peak1+2	38379	30	80768	210307
31peak1+3	17871	45	55177	222886
31peak1+4	15624	63	55177	234508
31peak1+5	17871	47	55177	220162

31peak2	7906	7	11508	28533
31peak2+3	8453	18	16990	66601
31peak2+4	11508	30	20403	82173
31peak2+5	6332	25	11508	66005
31peak3	16823	11	16823	39951
31peak3+4	42495	23	42495	75864
31peak3+5	7698	27	16823	74704
31peak4	5531	25	14160	51250
31peak4+5	37617	4	37617	66661
31peak5	2962	20	6767	38432
31	26208	37	80678	291908
32peak1	15286	26	21263	141286
32peak1+2	17061	29	31754	208886
32peak1+3	15286	46	43371	231819
32peak2	13146	12	28201	77716
32peak2+3	80221	8	80221	155960
32peak3	50230	6	50230	89848
32	17035	43	32899	289347
33	39659	1	39659	39787
34peak1	61597	7	61597	70020
34peak2	37431	1	37431	37559
34	61597	7	61597	104781
35peak1	28732	19	38239	164104
35peak1+2	17847	38	38239	286862
35peak1+3	28732	28	38239	181665
35peak2	14428	22	37676	128934
35peak2+3	15859	21	37676	131725
35peak3	2070	10	5715	19540
35	17035	38	38239	290414
36	37242	8	37654	89606
37peak1	4498	7	5903	20030
37peak2	40393	1	40393	40964

37	8330	12	14034	56965
38	3384	32	8039	71621
39peak1	4774	6	5936	21603
39peak1+2	2129	24	7130	45526
39peak1+3	28349	4	28349	43037
39peak2	17122	9	17122	29175
39peak2+3	30847	2	30847	42769
39peak3	4902	9	11092	28111
39	2936	29	7842	89326
40peak1	4610	8	5904	23196
40peak1+2	2368	21	10151	42327
40peak1+3	28323	6	28323	44714
40peak2	5098	10	10029	25956
40peak2+3	24440	2	24440	40985
40peak3	4899	8	11138	27737
40	2214	42	7645	92676
41peak1	9377	19	28649	73775
41peak1+2	9377	29	28649	123195
41peak1+3	7298	39	28098	120965
41peak1+4	10866	21	29001	124317
41peak2	6966	23	15723	56968
41peak2+3	15723	18	37615	89288
41peak2+4	4719	42	17164	109698
41peak3	4123	25	7298	49960
41peak3+4	52755	15	52755	89102
41peak4	46474	4	46474	52935
41	12686	34	28104	197493
42peak1	37635	13	40337	141857
42peak1+2	37635	12	40337	169657
42peak1+3	26313	45	40337	251343
42peak1+4	28864	21	40337	182167
42peak2	12539	8	13287	36575

42peak2+3	26020	33	44107	137844
42peak2+4	12539	16	18539	75114
42peak3	15695	32	38416	110098
42peak3+4	17023	27	49157	131383
42peak4	6093	9	13133	40574
42	28149	32	41380	290100
43peak1	4569	9	5926	23771
43peak1+2	2238	24	6849	49779
43peak1+3	8345	9	14961	43971
43peak2	17288	9	17288	29890
43peak2+3	30377	4	30377	41948
43peak3	3725	10	4965	25275
43	3853	21	6849	59799
44	52658	2	52658	90273
45	32297	16	38297	59980
46	10760	17	36835	78175
47peak1	13846	22	38353	131508
47peak1+2	13267	46	38353	185794
47peak1+3	13334	41	38353	163751
47peak2	5825	26	27811	65926
47peak2+3	27811	8	31014	77734
47peak3	1725	20	7470	32329
47	15903	29	38353	197360
48	3385	16	5970	60370
49	4467	5	4467	9004
50peak1	8869	30	22596	91228
50peak2	6170	16	12180	50722
50peak3	1572	13	9622	25190
50peak4	12610	3	16744	37545
50	17035	30	48272	169937
51peak1	28743	41	53682	240420
51peak2	7557	7	12657	37609

51	26325	26	53682	261433
52peak1	4647	7	10804	23951
52peak1+2	3434	20	8477	48690
52peak1+3	24884	8	24884	45133
52peak2	5146	10	10182	29905
52peak2+3	24366	2	24366	40880
52peak3	4953	9	25337	26428
52	8363	13	14150	58189
53	50013	1	50013	50141
54peak1	13518	27	21563	76551
54peak1+2	28260	24	40454	173159
54peak1+3	7554	40	21563	125819
54peak1+4	39613	30	42421	166594
54peak2	8036	35	18212	116717
54peak2+3	10790	32	18212	152946
54peak2+4	10780	37	35792	204704
54peak3	7296	15	15638	46534
54peak3+4	15785	24	37650	129745
54peak4	39613	4	42387	90353
54	21563	37	41593	290686
55peak1	37538	3	37538	41609
55peak1+2	37239	7	37239	62401
55peak1+3	37239	9	37239	46280
55peak1+4	37239	10	37239	71533
55peak1+5	37239	9	37239	58253
55peak2	10336	6	10336	21856
55peak2+3	4598	9	10336	24308
55peak2+4	3808	16	8167	49345
55peak2+5	6726	9	16320	37308
55peak3	563	5	841	5377
55peak3+4	17963	8	17963	33486
55peak3+5	3731	10	6062	24302

55peak4	15960	8	15960	33376
55peak4+5	24008	2	24008	43076
55peak5	5116	6	5966	19674
55	14052	14	37359	95804
56peak1	4824	32	38328	110855
56peak1+2	17771	53	38328	227332
56peak1+3	6917	48	38328	197121
56peak2	7242	40	28683	130343
56peak2+3	15911	38	28135	194146
56peak3	8016	19	29070	87161
56	17035	49	38328	290240
57	4085	2	4085	5429
58	3032	8	3249	12027
59	61772	1	61772	61900
60	9128	6	9128	13645
61peak1	4563	7	5912	20210
61peak2	10502	6	13506	43032
61	8089	14	25388	57594
62peak1	40441	17	40441	57718
62peak2	7046	5	13118	37174
62	40441	2	40441	77821
63	37715	1	37715	37843
64	61772	1	61772	61900
65peak1	4794	7	5940	23286
65peak1+2	2370	19	8434	44735
65peak1+3	28670	6	28670	45422
65peak2	5107	7	11109	25898
65peak2+3	39897	3	39897	42147
65peak3	4433	10	11138	28552
65	1193	57	5705	149586
66				
67	886	6	3110	8737

68peak1	4634	7	5897	21680
68peak1+2	2440	24	8823	45334
68peak1+3	11323	8	15440	44113
68peak2	17171	10	17171	29676
68peak2+3	30277	3	30277	42433
68peak3	4818		5867	27243
68	8089	17	24425	56387
69peak1	5752	36	14508	107985
69peak1+2	6786	29	31707	129600
69peak1+3	6070	46	14508	143523
69peak1+4	11709	40	37978	175623
69peak2	9168	9	12155	31610
69peak2+3	12155	13	16950	59817
69peak2+4	20588	13	37636	99504
69peak3	8547	9	10846	33016
69peak3+4	46492	4	46492	90513
69peak4	37617	5	37617	68206
69	31654	26	90429	214734
70peak1	1868	10	3712	12613
70peak2	514	1	514	917
70	3712	7	4114	12387
71	594	1	594	594
72	25633	2	25633	37385
73peak1	4800	6	5911	22640
73peak1+2	3831	19	8416	45177
73peak1+3	26778	7	26778	43629
73peak2	16442	10	16442	27598
73peak2+3	2292	28	5696	98725
73peak3	4943	8	11122	27622
73	2257	41	6582	121699
74peak1	10348	5	10348	21888
74peak1+2	2909	19	8954	45239

74peak1+3	26370	7	26370	42982
74peak2	15185	9	15185	27110
74peak2+3	40490	1	40490	41835
74peak3	4798	8	11054	27231
74	8237	11	14045	58867
75	37660	11	52654	151631
76peak1	15215	22	33462	119172
76peak2	611	1	611	611
76	19009	21	35341	119177
77	61772	1	61772	61772
78peak1	4612	9	5938	23816
78peak1+2	2397	20	7458	42756
78peak1+3	22565	7	22565	44473
78peak2	5150	10	10106	26350
78peak2+3	1293	38	2758	114250
78peak3	4923	8	11118	27496
78	1806	45	6512	111589
79	61772	1	61772	61900
80peak1	4616	7	10278	22703
80peak1+2	3845	17	10826	44548
80peak1+3	11346	6	20395	43298
80peak2	6007	9	10178	27975
80peak2+3	24423	2	24423	41335
80peak3	4910	11	5934	27160
80	8122	13	25526	56700
81	649	2	649	1249
82	2757	6	4408	12082
83peak1	2139	23	3881	36550
83peak2	1411	5	2167	6419
83peak3	535	3	750	3097
83peak4	697	2	697	1229
83	2747	27	5652	46287

84	28144	4	28144	38573
85peak1	10388	34	28289	143744
85peak1+2	12696	28	33837	170043
85peak1+3	9314	54	28150	210173
85peak1+4	11638	48	29041	219307
85peak1+5	10388	43	28289	158635
85peak1+6	10388	34	28289	144650
85peak2	2194	21	5611	36445
85peak2+3	7005	27	38471	90801
85peak2+4	8378	34	18408	108924
85peak2+5	2194	30	9622	55880
85peak2+6	2194	21	5611	36870
85peak3	8146	20	15463	65754
85peak3+4	12642	23	18408	126423
85peak3+5	12612	19	16849	84503
85peak3+6	8146	21	15463	66710
85peak4	13031	15	16480	73086
85peak4+5	17172	8	25826	77232
85peak4+6	13031	14	17172	73003
85peak5	1793	9	4861	14491
85peak5+6	1793	9	4861	14495
85peak6				
85	12853	43	34949	288310
86peak1	10314	5	10314	21027
86peak1+2	3916	16	11233	43014
86peak1+3	39510	3	39510	43615
86peak2	5738	8	11200	26244
86peak2+3	40661	1	40661	42022
86peak3	5017	8	11123	28234
86	8363	10	27106	58170
87	1608	17	2327	26840
88peak1	28218	7	38158	109767

88peak1+2	27799	30	38158	145979
88peak1+3	19162	27	38158	218148
88peak1+4	28733	13	38158	125841
88peak2	1210	23	12624	37558
88peak2+3	17138	21	37686	126684
88peak2+4	6464	25	16876	53294
88peak4	3539	7	5819	17619
88	26228	32	38158	239319
89peak1	11049	6	11049	23236
89peak1+2	2790	20	8463	46822
89peak1+3	19891	6	19988	44482
89peak2	15346	10	15346	28030
89peak2+3	24558	4	24558	42006
89peak3	4109	11	6154	28429
89	8363	13	22288	59100
90				
91	61772	1	61772	61772
92	61772	1	61772	61900
93peak1	4682	6	5901	19930
93peak2	12738	3	19454	40888
93	10045	10	15044	58139
94	882	3	1348	2800

Table 3.1: This table shows the SPAdes v3.1.1 assemblies that were done using the individual and multiple peaks extracted from the K-mer graphs (Appendix A). The highlighted raw were used for as the best representation of the sample. The table shows the N50, number of contigs >=500bp, largest contig (bp) and the total length >= 0 bp in each assembly.

Appendix 3 *Pseudomonas aeruginosa* phage genomes

The temperate phage vB_Pae_CF1a genome size was 31.8 kb, the phage was assembled using SPAdes and extended using PriceTI as the SPAdes assemblies had the phage in separate contigs. The genome map is shown in Figure 1. The total number of genes identified was 48 of which 11 were identified with putative functions Table 1. Blastn on Viruses only database showed an 8 kb / 8.6 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF1a had a 62.5 % G-C content. No tRNA gene was identified and no integrase was found.

Start	End	Direction	Putative functional protein
913	527	Reverse	Bacterial regulatory proteins, luxR family
2787	1996	Reverse	HTH-type transcriptional regulator PrtR
10510	10899	Forward	Phage antitermination protein Q
11438	11965	Forward	Phage regulatory protein Rha (Phage_pRha)
12050	12382	Forward	Phage holin family (Lysis protein S)
12379	12996	Forward	Chitinase class I
13232	13702	Forward	Bacteriophage lysis protein
14559	15104	Forward	Phage DNA packaging protein Nu1
15076	17040	Forward	Phage terminase large subunit (GpA)
17246	18892	Forward	Phage portal protein, lambda family
18864	20945	Forward	ATP-dependent Clp protease proteolytic subunit

Table 1: Functional genes identified for vB_Pae_CF1a.

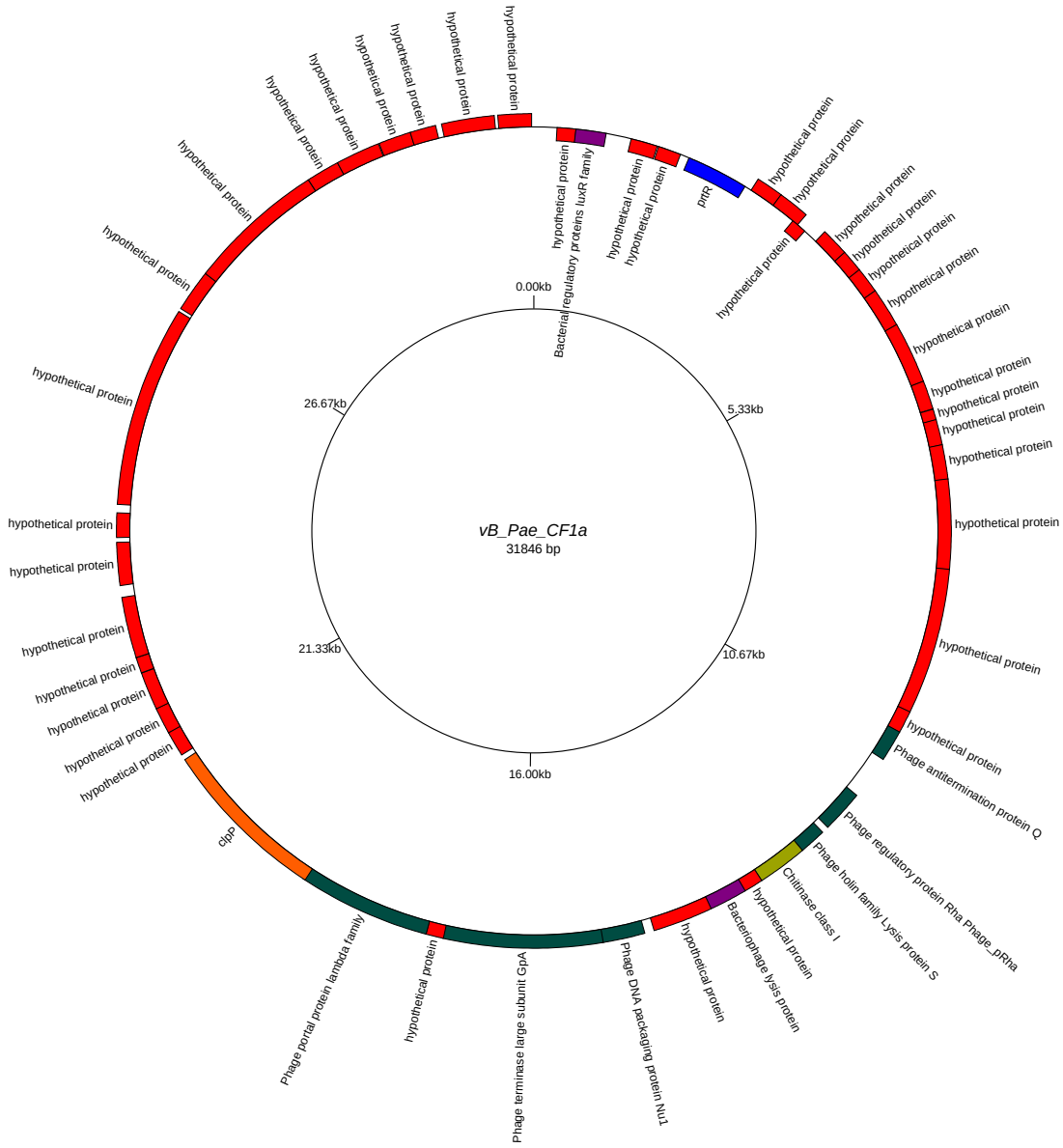


Figure 1: Genome map of vB_Pae_CF1a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF2a genome size was 36.3 kb, the phage was assembled using SPAdes all the reads were used and gave a mean coverage of 5.5. The genome map is shown in Figure 2. The total number of genes identified was 48 of which 7 were identified with putative functions Table 3. Blastn on viruses only database showed a 9.6 kb / 10.5 kb match with *Pseudomonas* phage B3 (GenBank: NC_006548.1). The phage vB_Pae_CF2a had a 62.9 % G-C content. No tRNA gene was identified and no integrase was found however, Mu like prophage proteins are identified.

Start	**End**	**Direction**	**Putative functional protein**
8117	8506	Forward	Helix-turn-helix domain protein
10378	11007	Forward	Soluble lytic murein transglycosylase precursor
13004	14677	Forward	Terminase-like family protein
16235	17473	Forward	Phage Mu protein F like protein
17475	18050	Forward	Phage virion morphogenesis family protein
18261	19370	Forward	Mu-like prophage I protein
19795	20691	Forward	Mu-like prophage major head subunit gpT

Table 2: Functional genes identified for vB_Pae_CF2a.

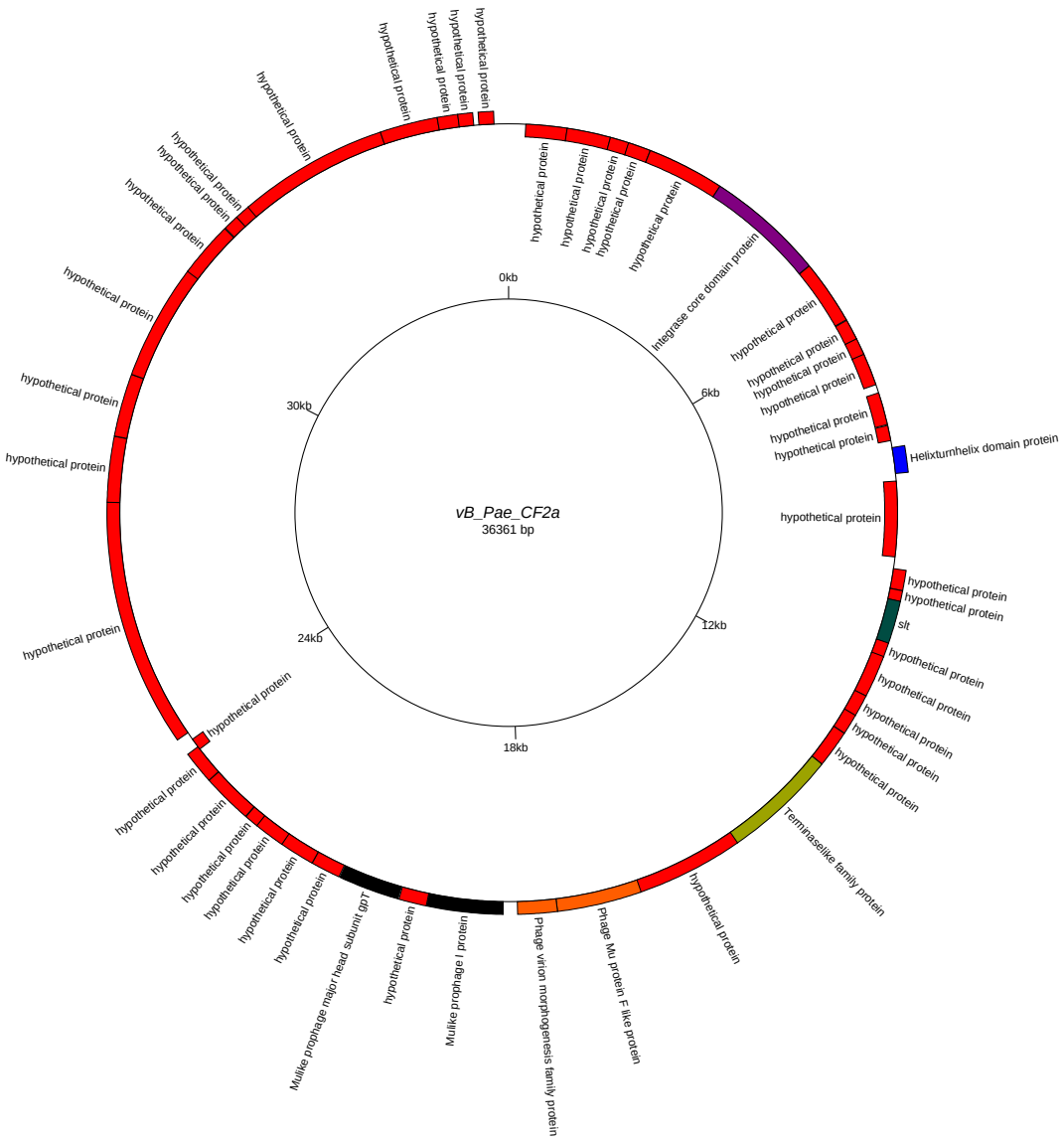
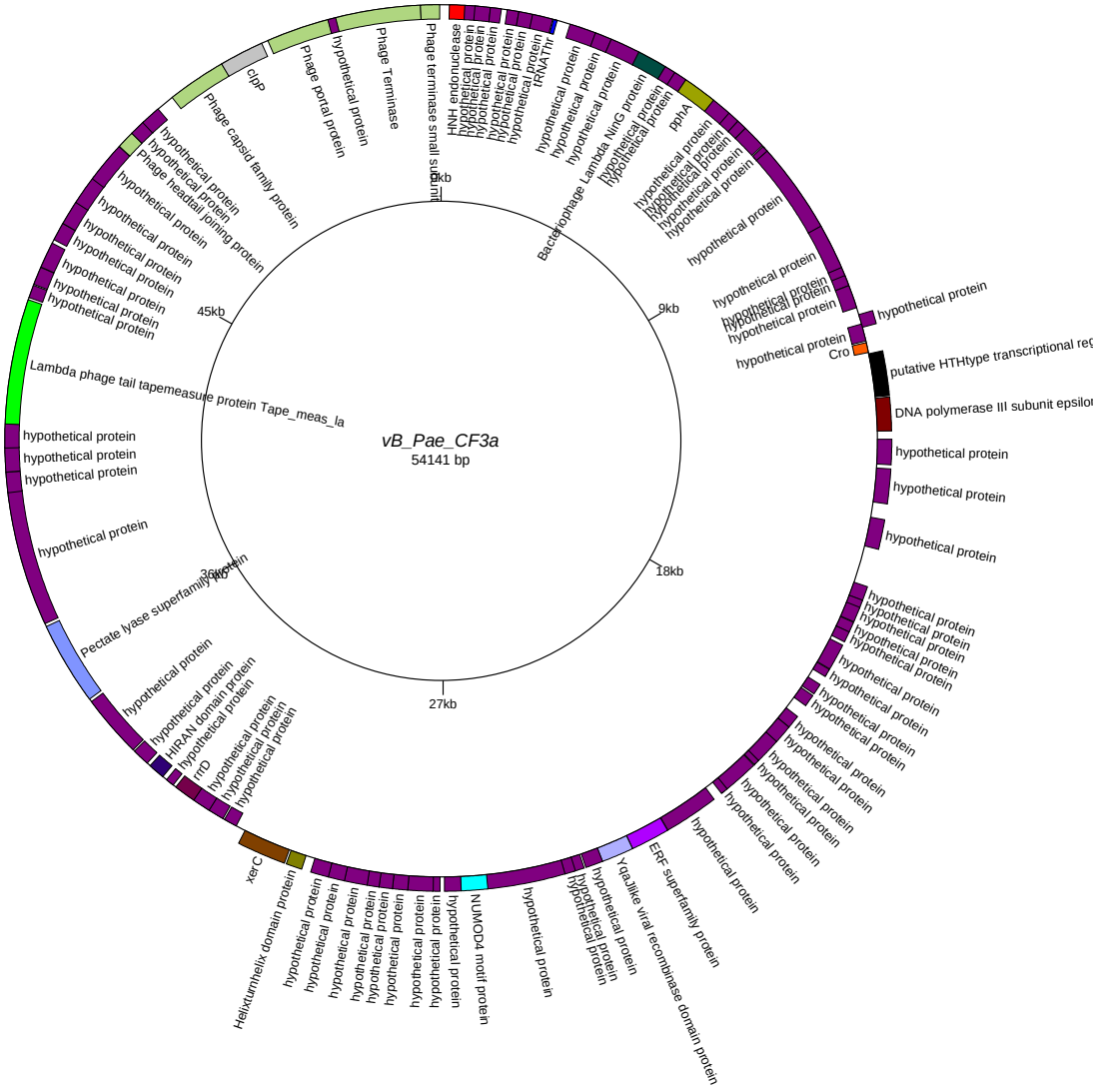


Figure 2: Genome map of vB_Pae_CF2a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF3a genome size was 54.1 kb, the phage was assembled using SPAdes all the reads were used and gave a mean coverage of 21. The genome map of the phage is shown in Figure 3. The total number of genes identified was 97 of which 22 were identified with putative functions Table 3. Blastn on viruses only database showed a 6.3 kb / 7 kb match with *Pseudomonas* phage vB_PaeS_PMG1 (GenBank: NC_016765.1). The phage vB_Pae_CF3a had a 58.4 % G-C content. A tRNA gene was identified along with structural and lysogenic genes.

Start	**End**	**Direction**	**Putative functional protein**
476	162	Reverse	HNH endonuclease
2320	2245	Reverse	tRNA-Thr(tgt)
4652	4053	Reverse	Bacteriophage Lambda NinG protein
5816	5130	Reverse	Serine/threonine-protein phosphatase 1
11776	11576	Reverse	Cro
11801	12664	Forward	putative HTH-type transcriptional regulator
12694	13344	Forward	DNA polymerase III subunit epsilon
22534	23286	Forward	ERF superfamily protein
23290	23916	Forward	YqaJ-like viral recombinase domain protein
26174	26689	Forward	NUMOD4 motif protein
29780	30094	Forward	Helix-turn-helix domain protein
30133	31089	Forward	Tyrosine recombinase XerC
32700	32266	Reverse	Lysozyme RrrD
33363	33031	Reverse	HIRAN domain protein
36816	35146	Reverse	Pectate lyase superfamily protein
43445	40941	Reverse	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
47351	46995	Reverse	Phage head-tail joining protein
49603	48416	Reverse	Phage capsid family protein
50490	49600	Reverse	ATP-dependent Clp protease proteolytic subunit
51884	50622	Reverse	Phage portal protein
53729	52038	Reverse	Phage Terminase
54111	53731	Reverse	Phage terminase, small subunit

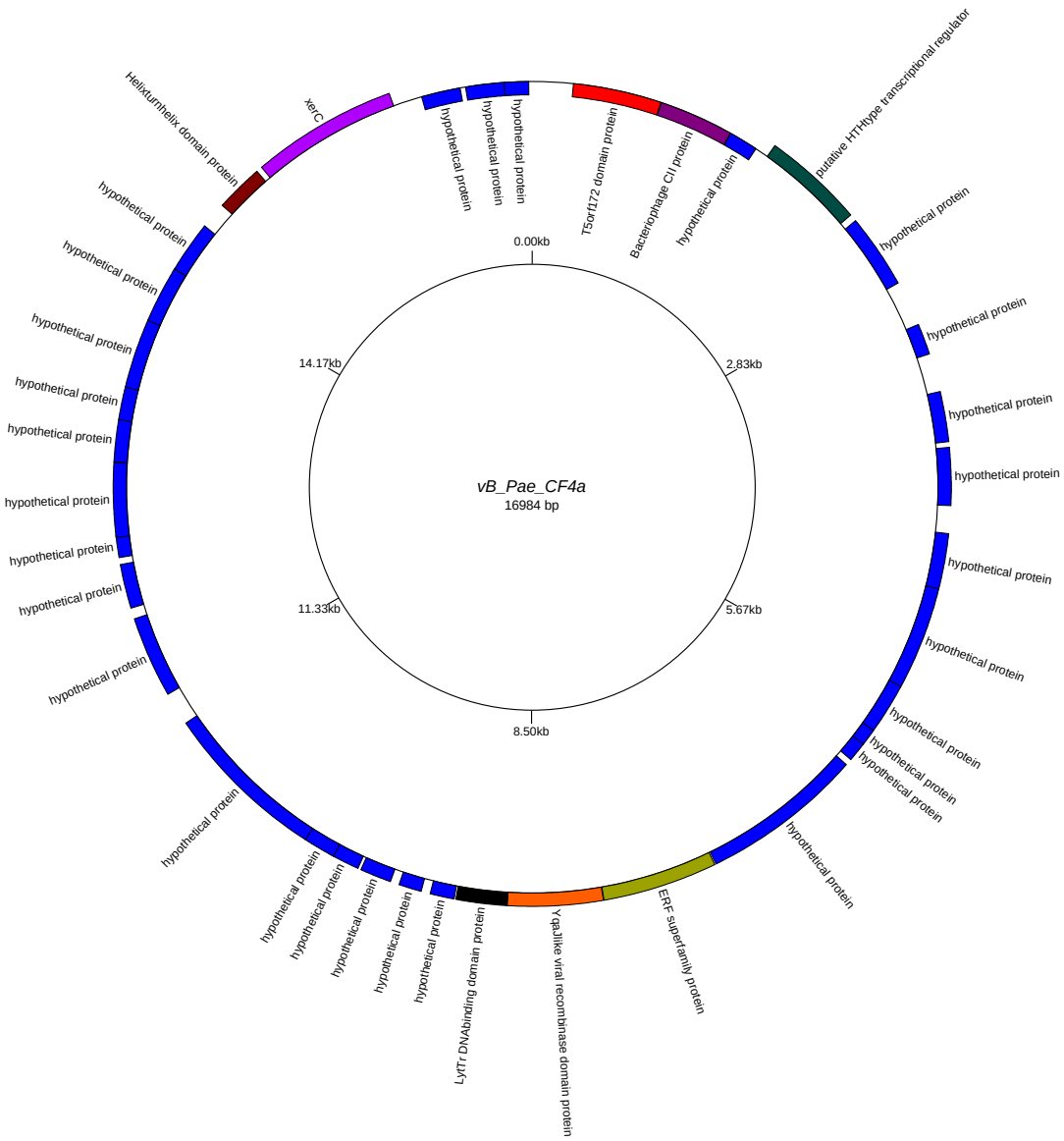
Table 3: Functional genes identified for vB_Pae_CF3a.


vB_Pae_CF3a
54141 bp
0kb
9kb
18kb
27kb
36kb
45kb
Phage terminase small subunit
HNH endonuclease
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
tRNA-Thr
hypothetical protein
hypothetical protein
Bacteriophage Lambda NinG protein
hypothetical protein
pphA
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Cro
putative HTHtype transcriptional reg
DNA polymerase III subunit epsilon
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein

The incomplete temperate phage vB_Pae_CF4a genome size was 16.9 kb, the phage was assembled using SPAdes all the reads were used and gave a mean coverage of 21. This is possibly an incomplete phage more similar to Phi297. The genome map of this phage is shown in Figure 4. The total number of genes identified was 37 of which 8 were identified with putative functions these can be seen in Table 4. Blastn on viruses only database showed a 3.3 kb / 3.6 kb match with *Pseudomonas* phage D3 (GenBank: NC_002484.1). The phage vB_Pae_CF4a had a 58.1 % G-C content. No tRNA gene was identified however; a lysogenic gene encoding CII was identified.

Start	**End**	**Direction**	**Putative functional protein**
872	276	Reverse	T5orf172 domain protein
1379	876	Reverse	Bacteriophage CII protein
1669	2334	Forward	putative HTH-type transcriptional regulator
7274	8026	Forward	ERF superfamily protein
8030	8656	Forward	YqaJ-like viral recombinase domain protein
8653	8988	Forward	LytTr DNA-binding domain protein
14733	15047	Forward	Helix-turn-helix domain protein
15086	16042	Forward	Tyrosine recombinase XerC

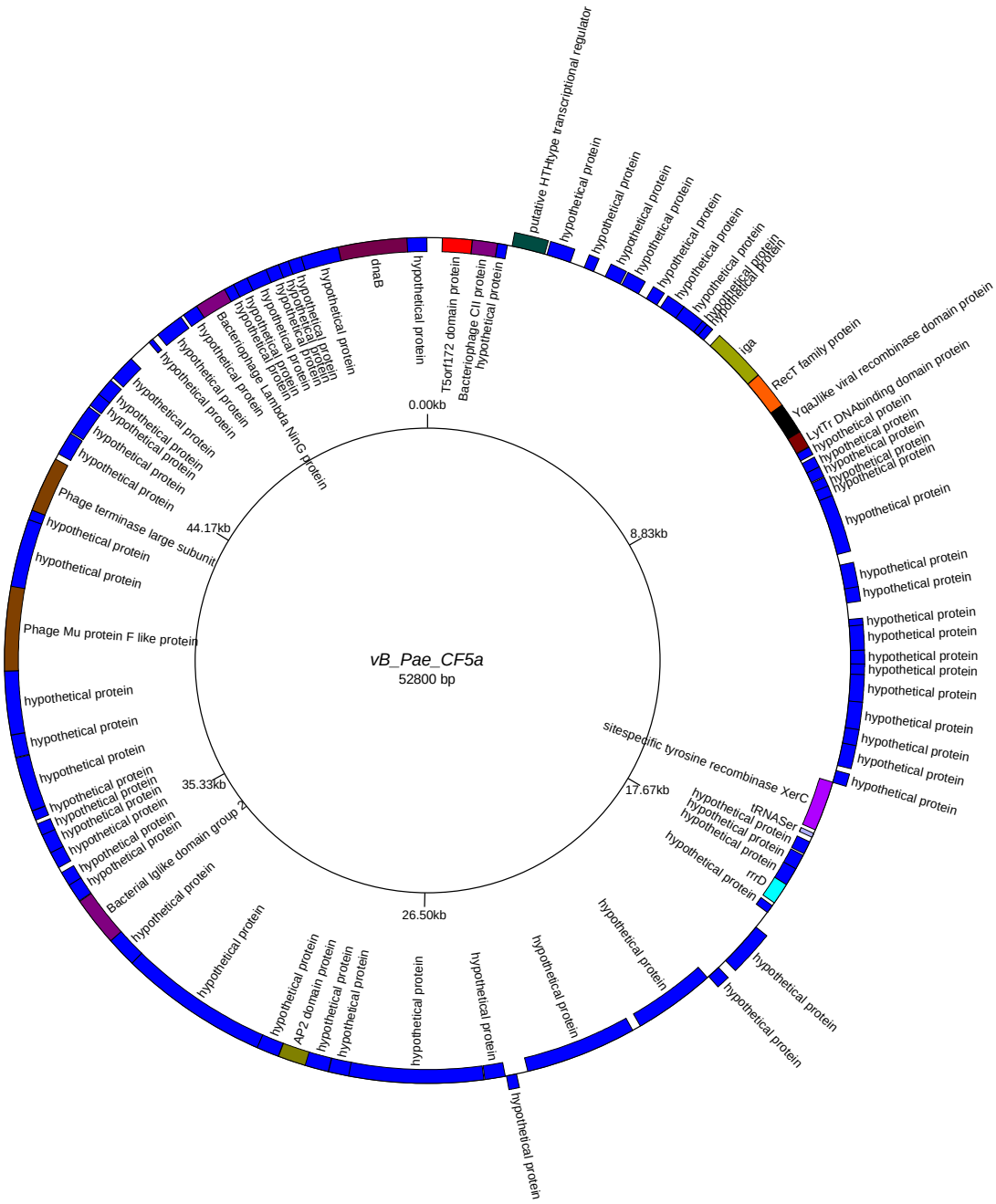
Table 4: Functional genes identified for vB_Pae_CF4a.


vB_Pae_CF4a
16984 bp
0.00kb
2.83kb
5.67kb
8.50kb
11.33kb
14.17kb
hypothetical protein
hypothetical protein
hypothetical protein
T5orf172 domain protein
Bacteriophage CII protein
hypothetical protein
putative HTHtype transcriptional regulator
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Pae_CF5a genome size was 52.8 kb, the phage was assembled using SPAdes and has a mean coverage of 315. The genome map of this phage is shown in Figure 5. The total number of genes identified was 86 of which 18 were identified with putative functions these can be seen in Table 5. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_CF5a had a 58.9 % G-C content. A tRNA gene, CII and Ig-like domains were all identified along with recombinase genes.

Start	End	Direction	Putative functional protein
891	295	Reverse	T5orf172 domain protein
1398	895	Reverse	Bacteriophage CII protein
1688	2353	Forward	putative HTH-type transcriptional regulator
6147	7211	Forward	IgA-specific serine endopeptidase autotransporter precursor
7219	7965	Forward	RecT family protein
7949	8569	Forward	YqaJ-like viral recombinase domain protein
8566	8901	Forward	LytTr DNA-binding domain protein
16663	15653	Reverse	site-specific tyrosine recombinase XerC
16824	16737	Reverse	tRNA-Ser(cga)
18328	17894	Forward	Lysozyme RrrD
18931	19830	Forward	BRO family, N-terminal domain
21971	20325	Reverse	D-glucuronyl C5-epimerase C-terminus
29426	28866	Reverse	AP2 domain protein
34507	33512	Reverse	Bacterial Ig-like domain (group 2)
41122	39392	Reverse	Phage Mu protein F like protein
43773	42655	Reverse	Phage terminase large subunit
48598	47951	Reverse	Bacteriophage Lambda NinG protein
52397	51012	Reverse	Replicative DNA helicase

Table 5: Functional genes identified for vB_Pae_CF5a.


vB_Pae_CF5a
52800 bp
0.00kb
8.83kb
17.67kb
26.50kb
35.33kb
44.17kb
putative HTHtype transcriptional regulator
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
tga
RecT family protein
YqaJlike viral recombinase domain protein
LytTr DNAbinding domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
sitespecific tyrosine recombinase XerC
tRNASer
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Pae_CF5b genome size was 37.7 kb, the phage was assembled using SPAdes and has a mean coverage of 201. The genome map of this phage is shown in Figure 6. The total number of genes identified was 58 of which 10 were identified with putative functions these can be seen in Table 6. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF5b had a 64.2 % G-C content. No tRNA gene was identified however; Mu-like phage structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
1261	548	Reverse	HTH-type transcriptional regulator PrtR
1431	1787	Forward	DNA-binding transcriptional regulator Nlp
2489	4483	Forward	Mu DNA-binding domain protein
6696	7214	Forward	Bacteriophage Mu Gam like protein
10018	10467	Forward	Mor transcription activator family protein
16936	18222	Forward	Phage Mu protein F like protein
18222	18689	Forward	Phage virion morphogenesis family protein
19708	20814	Forward	Mu-like prophage I protein
21250	22158	Forward	Mu-like prophage major head subunit gpT
25190	28750	Forward	Prophage tail length tape measure protein

Table 6: Functional genes identified for vB_Pae_CF5b.

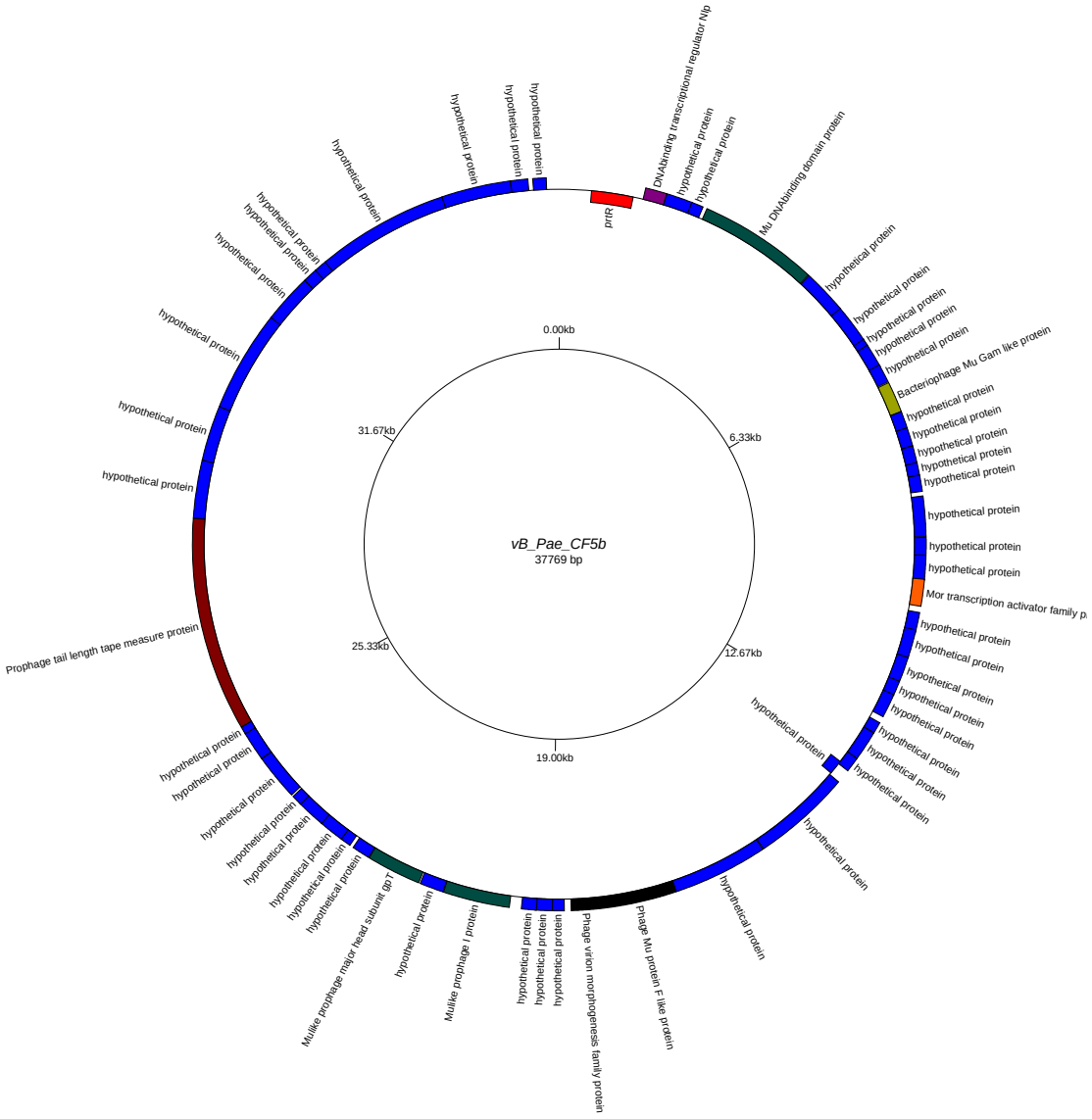


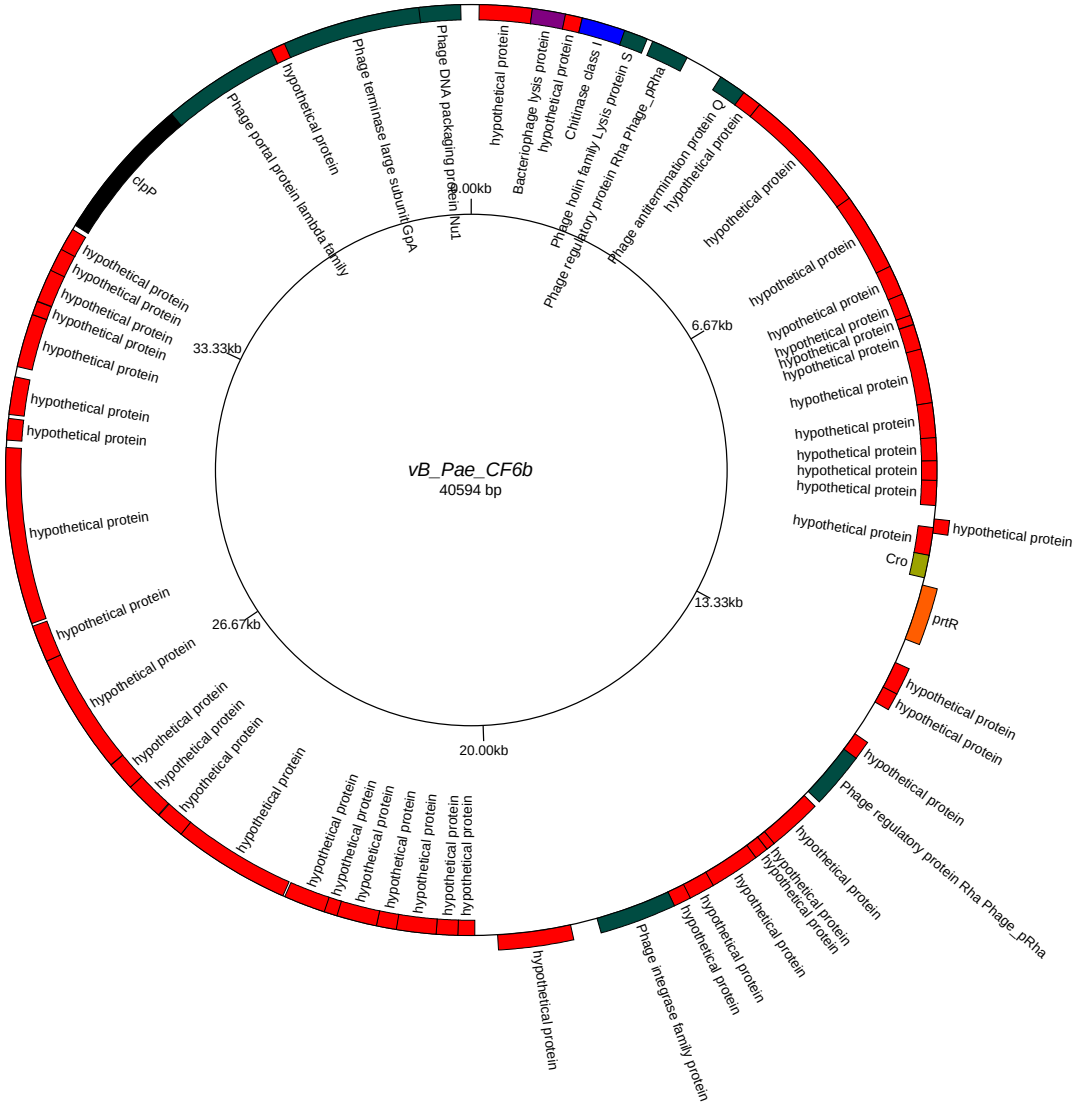
Figure 6: Genome map of vB_Pae_CF5b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF6a genome size was 61.7 kb, the phage was assembled using SPAdes and has a mean coverage of 86. The genome map of this phage is shown in Figure 7. The total number of genes identified was 62 of which 14 were identified with putative functions these can be seen in Table 7. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66

(GenBank: KC262634.1). The phage vB_Pae_CF6a had a 63.7 % G-C content. No tRNA gene was identified however; a CPS-53 integrase like gene was identified.

Start	End	Direction	Putative functional protein
4605	5102	Forward	Single-stranded DNA-binding protein
5129	6349	Forward	Recombination-associated protein RdgC
6346	6555	Forward	LexA repressor
6700	9330	Forward	DNA methylase
9327	11135	Forward	C-5 cytosine-specific DNA methylase
13664	13897	Forward	Response regulator inhibitor for tor operon
15133	13898	Reverse	Putative prophage CPS-53 integrase
18738	18199	Reverse	Phage lysozyme
47162	46290	Reverse	RyR domain protein
53961	52678	Reverse	Phage terminase large subunit
55759	55145	Reverse	Bacteriophage Lambda NinG protein
58365	58165	Reverse	Cro
58473	59270	Forward	HTH-type transcriptional regulator PrtR
60689	61060	Forward	Carbon storage regulator homolog

Table 7: Functional genes identified for vB_Pae_CF6a.


vB_Pae_CF6b
40594 bp
0.00kb
6.67kb
13.33kb
20.00kb
26.67kb
33.33kb
hypothetical protein
Bacteriophage lysis protein
hypothetical protein
Chitinase class I
Phage holin family Lysis protein S
Phage regulatory protein Rha Phage_pRha
phage antitermination protein Q
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Cro
prtR
hypothetical protein
hypothetical protein
hypothetical protein
Phage regulatory protein Rha Phage_pRha
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage integrase family protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Pae_CF6b genome size was 40.5 kb, the phage was assembled using SPAdes and has a mean coverage of 881. The genome map of this phage is shown in Figure 8. The total number of genes identified was 61 of which 13 were identified with putative functions these can be seen in Table 8. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF6b had a 61.5 % G-C content. No tRNA gene was identified however; a *Cro* gene and phage integrase like gene was identified.

Start	**End**	**Direction**	**Putative functional protein**
1320	850	Reverse	Bacteriophage lysis protein
2173	1556	Reverse	Chitinase class I
2502	2170	Reverse	Phage holin family (Lysis protein S)
3114	2587	Reverse	Phage regulatory protein Rha (Phage_pRha)
4042	3653	Reverse	Phage antitermination protein Q
11658	11338	Reverse	Cro
11761	12549	Forward	HTH-type transcriptional regulator PrtR
14303	15070	Forward	Phage regulatory protein Rha (Phage_pRha)
17479	18531	Forward	Phage integrase family protein
36103	34022	Reverse	ATP-dependent Clp protease proteolytic subunit
37721	36075	Reverse	Phage portal protein, lambda family
39891	37927	Reverse	Phage terminase large subunit (GpA)
40447	39863	Reverse	Phage DNA packaging protein Nu1

Table 8: Functional genes identified for vB_Pae_CF6b.

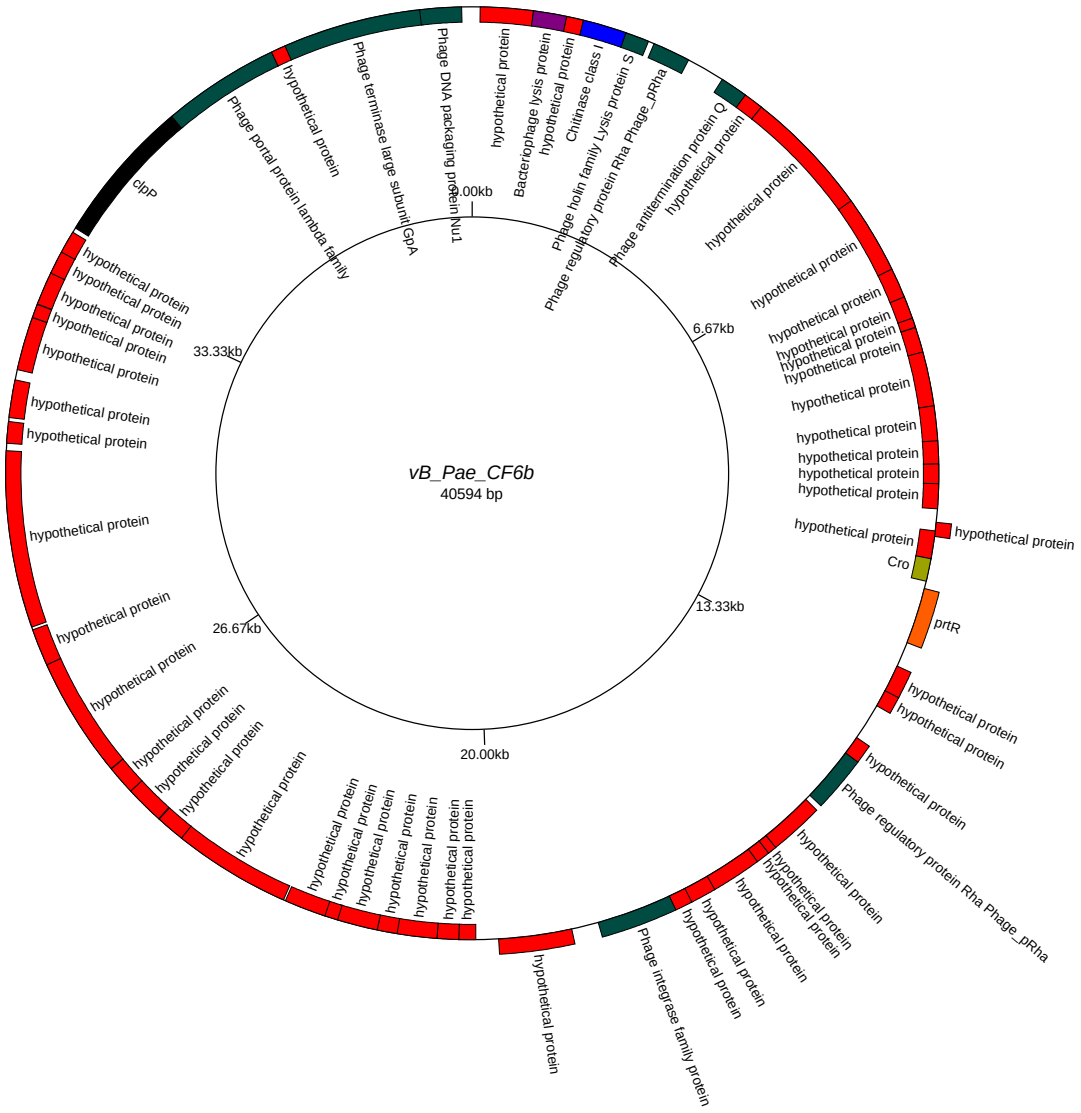


Figure 8: Genome map of vB_Pae_CF6b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF7a genome size was 37.3 kb, the phage was assembled using SPAdes and has a mean coverage of 154. The genome map of this phage is shown in Figure 9. The total number of genes identified was 56 of which 8 were identified with putative functions these can be seen in Table 9. Blastn on viruses only database showed a 27.1 kb / 27.8 kb match with *Pseudomonas* phage JBD24 (GenBank: NC_020203.1). The phage vB_Pae_CF7a had a 64.1 % G-C content. No tRNA gene was identified however; several Mu like proteins were identified associated with morphogenesis and structural proteins.

Start	End	Direction	Putative functional protein
889	230	Reverse	putative HTH-type transcriptional regulator
1130	1486	Forward	DNA-binding transcriptional regulator Nlp
2649	4718	Forward	Mu DNA-binding domain protein
10105	10554	Forward	Mor transcription activator family protein
17012	18298	Forward	Phage Mu protein F like protein
18298	18765	Forward	Phage virion morphogenesis family protein
19803	20900	Forward	Mu-like prophage I protein
20904	21818	Forward	Mu-like prophage major head subunit gpT

Table 9: Functional genes identified for vB_Pae_CF7a.

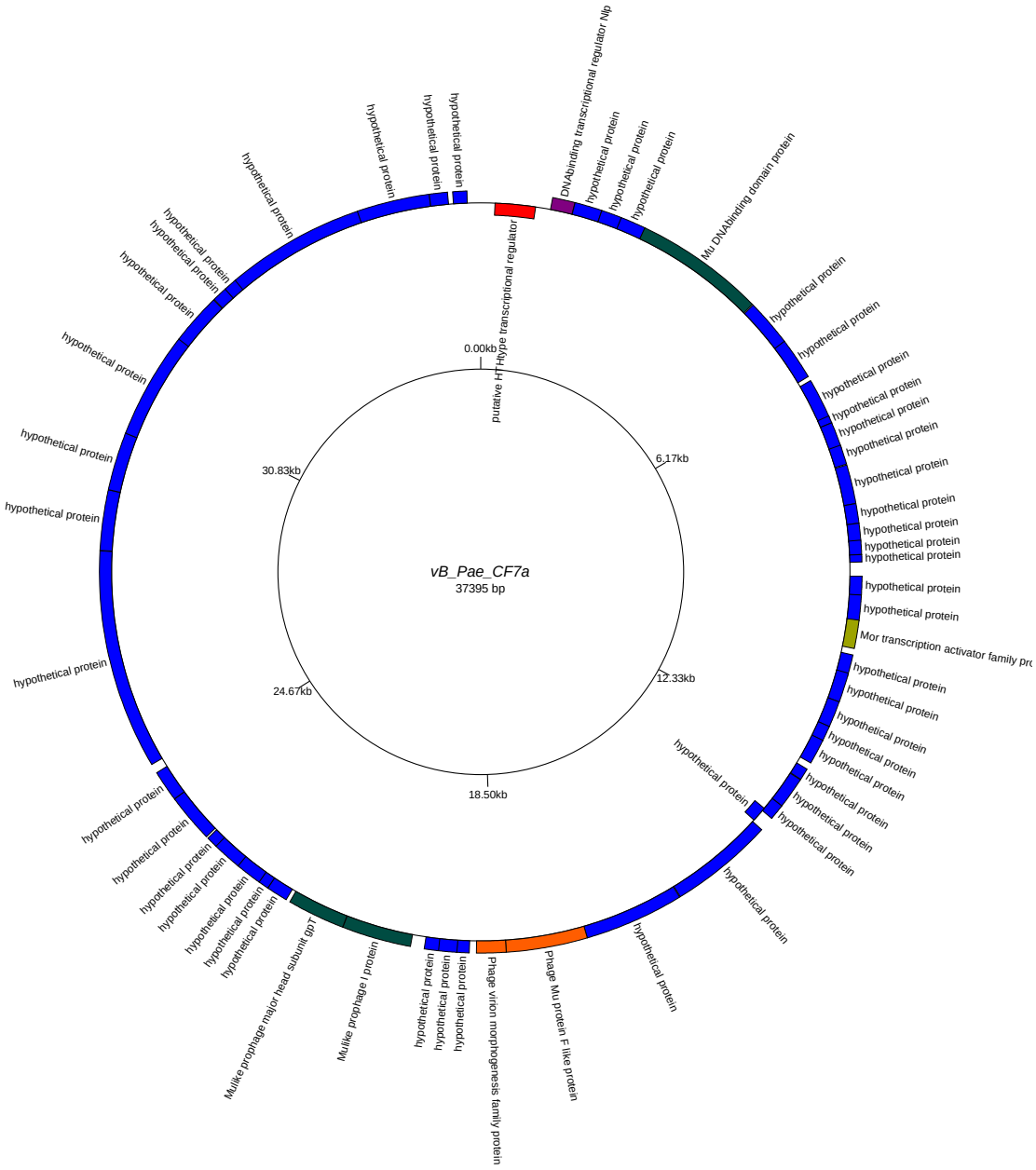


Figure 9: Genome map of vB_Pae_CF7a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF7b genome size was 46.4 kb, the phage was assembled using SPAdes and then extended using PriceTI. The genome map of this phage is shown in Figure 10. The total number of genes identified was 70 of which 17 were identified with putative functions these can be seen in Table 10. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF7b had a 61.3 % G-C content. No tRNA gene was identified however; several proteins were identified associated with lysogenic life cycle including integrase and a *Cro* gene.

Start	End	Direction	Putative functional protein
1227	997	Reverse	Arc-like DNA binding domain protein
2070	1288	Reverse	Phage regulatory protein Rha (Phage_pRha)
3475	3089	Reverse	Bacterial regulatory proteins, luxR family
5349	4558	Reverse	HTH-type transcriptional regulator PrtR
13072	13461	Forward	Phage antitermination protein Q
14000	14527	Forward	Phage regulatory protein Rha (Phage_pRha)
14612	14944	Forward	Phage holin family (Lysis protein S)
14941	15558	Forward	Chitinase class I
15794	16264	Forward	Bacteriophage lysis protein
17120	17665	Forward	Phage DNA packaging protein Nu1
17637	19601	Forward	Phage terminase large subunit (GpA)
19807	21453	Forward	Phage portal protein, lambda family
21425	23506	Forward	ATP-dependent Clp protease proteolytic subunit
40048	38996	Reverse	Phage integrase family protein
43224	42625	Reverse	Phage regulatory protein Rha (Phage_pRha)
45766	44978	Reverse	HTH-type transcriptional regulator PrtR
45869	45997	Forward	Cro

Table 10: Functional genes identified for vB_Pae_CF7b.

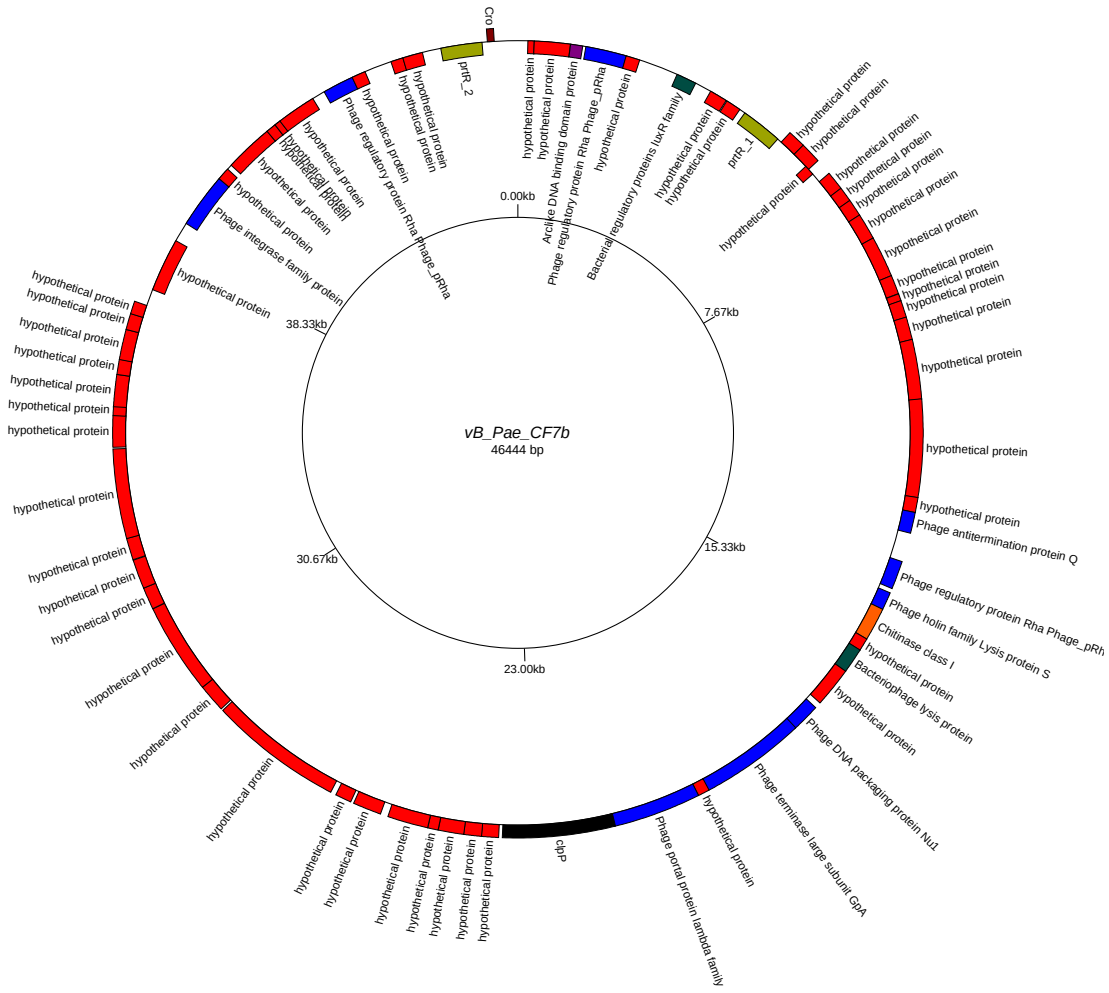


Figure 10: Genome map of vB_Pae_CF7b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF9a genome size was 41.6 kb, the phage was assembled using SPAdes and then extended using PriceTI. The genome map of this phage is shown in Figure 11. The total number of genes identified was 66 of which 13 were identified with putative functions these can be seen in Table 11. Blastn on viruses only database showed a 9.6 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF9a had a 61.5 % G-C content. No tRNA gene was identified however; several proteins were identified associated with structural proteins and lytic life cycle.

Start	End	Direction	Putative functional protein
443	57	Reverse	Bacterial regulatory proteins, luxR family
2317	1526	Reverse	HTH-type transcriptional regulator PrtR
10040	10429	Forward	Phage antitermination protein Q
10968	11495	Forward	Phage regulatory protein Rha (Phage_pRha)
11580	11912	Forward	Phage holin family (Lysis protein S)
11909	12526	Forward	Chitinase class I
12762	13232	Forward	Bacteriophage lysis protein
14089	14634	Forward	Phage DNA packaging protein Nu1
14606	16570	Forward	Phage terminase large subunit (GpA)
16776	18422	Forward	Phage portal protein, lambda family
18394	20505	Forward	ATP-dependent Clp protease proteolytic subunit
39430	39200	Reverse	Arc-like DNA binding domain protein
40277	39510	Reverse	Phage regulatory protein Rha (Phage_pRha)

Table 11: Functional genes identified for vB_Pae_CF9a.

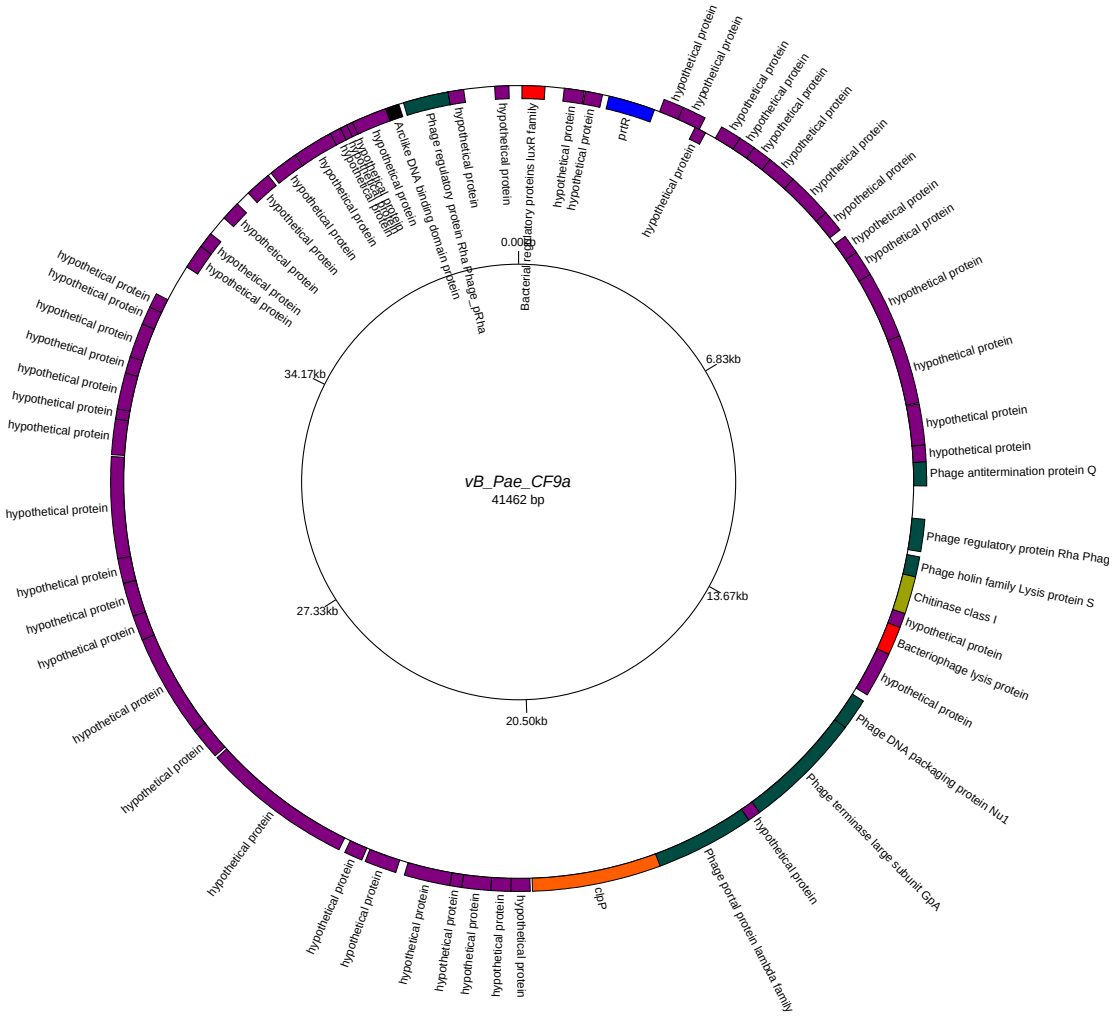


Figure 11: Genome map of vB_Pae_CF9a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF13a genome size was 54.1 kb, the phage was assembled using SPAdes and has a mean coverage of 15. The genome map of this phage is shown in Figure 12. The total number of genes identified was 97 of which 22 were identified with putative functions these can be seen in Table 12. Blastn on viruses only database showed a 6.3 kb / 7 kb match with *Pseudomonas* phage vB_PaeS_PMG1 (GenBank: NC_016765.1). The phage vB_Pae_CF13a had a 58.4 % G-C content. A tRNA gene was identified along with several proteins associated with structural proteins and lysogenic protein.

Start	End	Direction	Putative functional protein
474	160	Reverse	HNH endonuclease
2318	2243	Reverse	tRNA-Thr(tgt)
4650	4051	Reverse	Bacteriophage Lambda NinG protein
5814	5128	Reverse	Serine/threonine-protein phosphatase 1
11774	11574	Reverse	Cro
11799	12662	Forward	putative HTH-type transcriptional regulator
12692	13342	Forward	DNA polymerase III subunit epsilon
22532	23284	Forward	ERF superfamily protein
23288	23914	Forward	YqaJ-like viral recombinase domain protein
26172	26687	Forward	NUMOD4 motif protein
29778	30092	Forward	Helix-turn-helix domain protein
30131	31087	Forward	Tyrosine recombinase XerC
32698	32264	Reverse	Lysozyme RrrD
33361	33029	Reverse	HIRAN domain protein
36814	35144	Reverse	Pectate lyase superfamily protein
43443	40939	Reverse	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
47349	46993	Reverse	Phage head-tail joining protein
49601	48414	Reverse	Phage capsid family protein
50488	49598	Reverse	ATP-dependent Clp protease proteolytic subunit
51882	50620	Reverse	Phage portal protein
53727	52036	Reverse	Phage Terminase

54109	53729	Reverse	Phage terminase, small subunit

Table 12: Functional genes identified for vB_Pae_CF13a.

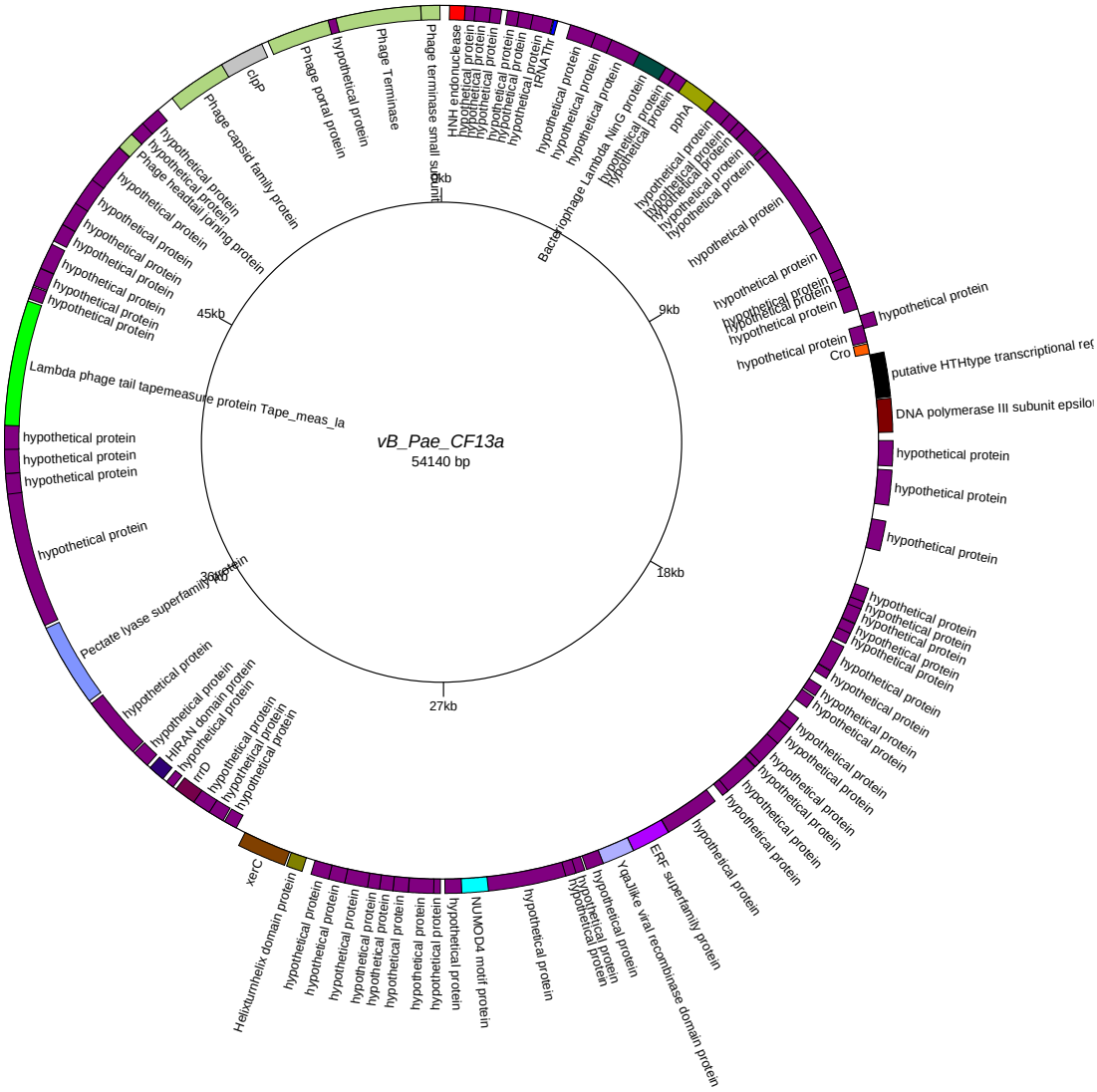
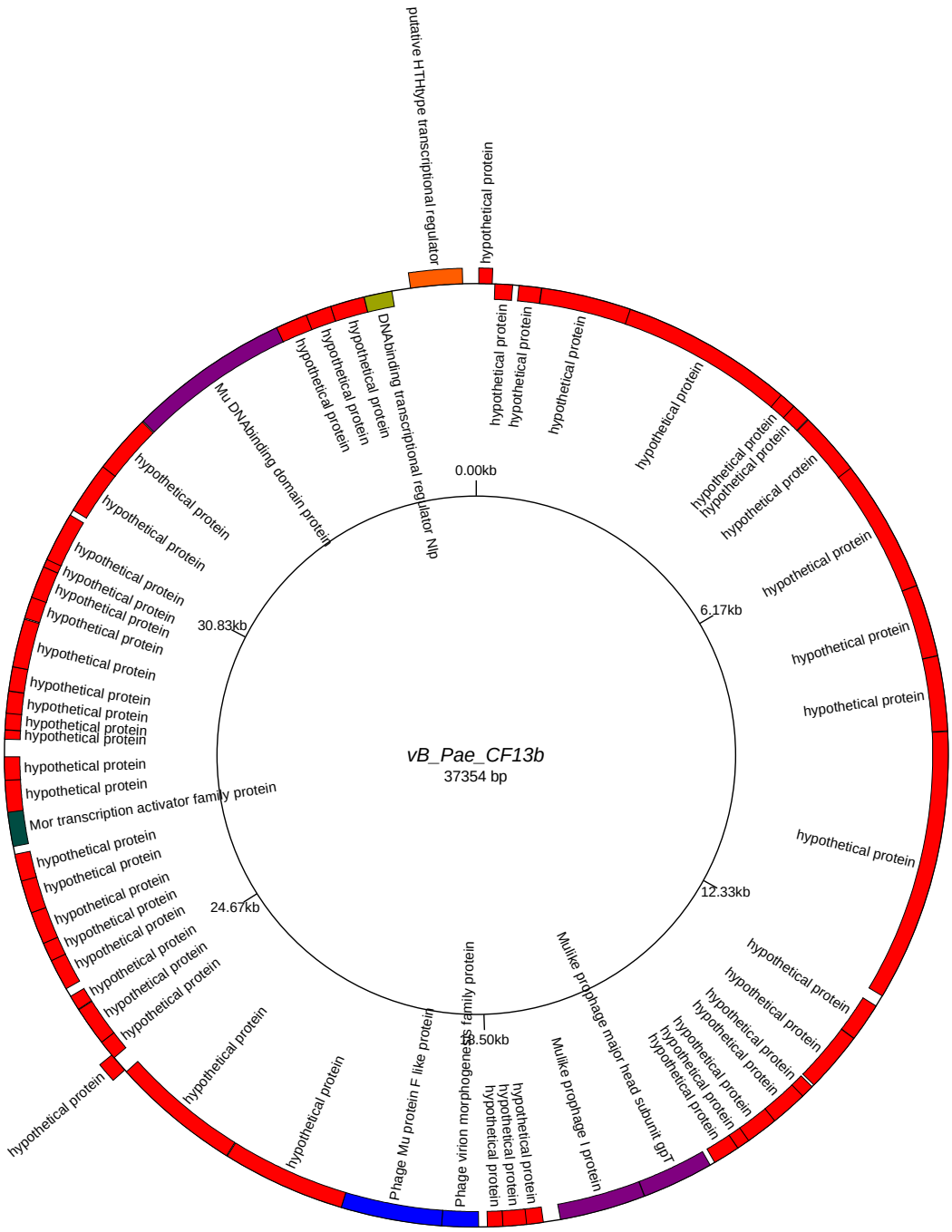


Figure 12: Genome map of vB_Pae_CF13a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF13b genome size was 37.3 kb, the phage was assembled using SPAdes and has a mean coverage of 12. The genome map of this phage is shown in Figure 13. The total number of genes identified was 57 of which 8 were identified with putative functions these can be seen in Table 13. Blastn on viruses only database showed a 27.1 kb / 27.8 kb match with *Pseudomonas* phage JBD24 (GenBank: NC_020203.1). The phage vB_Pae_CF13b had a 64.1 % G-C content. No tRNA gene was identified however; several genes associated with structural proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
16509	15595	Reverse	Mu-like prophage major head subunit gpT
17610	16513	Reverse	Mu-like prophage I protein
19115	18648	Reverse	Phage virion morphogenesis family protein
20401	19115	Reverse	Phage Mu protein F like protein
27308	26859	Reverse	Mor transcription activator family protein
34764	32695	Reverse	Mu DNA-binding domain protein
36283	35927	Reverse	DNA-binding transcriptional regulator Nlp
36524	37183	Forward	putative HTH-type transcriptional regulator

Table 13: Functional genes identified for vB_Pae_CF13b.


vB_Pae_CF13b
37354 bp
0.00kb
6.17kb
12.33kb
18.50kb
24.67kb
30.83kb
putative HTHtype transcriptional regulator
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Pae_CF14a genome size was 38.3 kb, the phage was assembled using SPAdes and has a mean coverage of 22. The genome map of this phage is shown in Figure 14. The total number of genes identified was 58 of which 26 were identified with putative functions these can be seen in Table 14. Blastn on viruses only database showed a 1.9 kb / 1.9 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF14a had a 61.3 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
718	389	Reverse	HNH endonuclease
1425	913	Reverse	Bacteriophage lysis protein
2042	1425	Reverse	Chitinase class I
2371	2039	Reverse	Phage holin family (Lysis protein S)
2986	2456	Reverse	Phage regulatory protein Rha (Phage_pRha)
4150	3761	Reverse	Phage antitermination protein Q
5819	4422	Reverse	Replicative DNA helicase
6631	5816	Reverse	DNA replication protein DnaC
10892	10587	Reverse	Helix-turn-helix
11002	11661	Forward	LexA repressor
11810	12196	Forward	Bacterial regulatory proteins, luxR family
12508	13293	Forward	BRO family, N-terminal domain
13368	13598	Forward	Arc-like DNA binding domain protein
18076	19059	Forward	Transposase DDE domain protein
23329	19664	Reverse	Carbohydrate binding domain protein
23958	23386	Reverse	Bacteriophage lambda tail assembly protein I
25467	24709	Reverse	NlpC/P60 family protein
26216	25470	Reverse	Phage minor tail protein L
26551	26213	Reverse	Phage minor tail protein
29826	26551	Reverse	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
32524	32198	Reverse	Phage head-tail joining protein

32847	32524	Reverse	Phage gp6-like head-tail connector protein
34316	33102	Reverse	Phage capsid family protein
34957	34313	Reverse	Caudovirus prohead protease
36164	34941	Reverse	Phage portal protein
37846	36167	Reverse	Phage Terminase

Table 14: Functional genes identified for vB_Pae_CF14a.

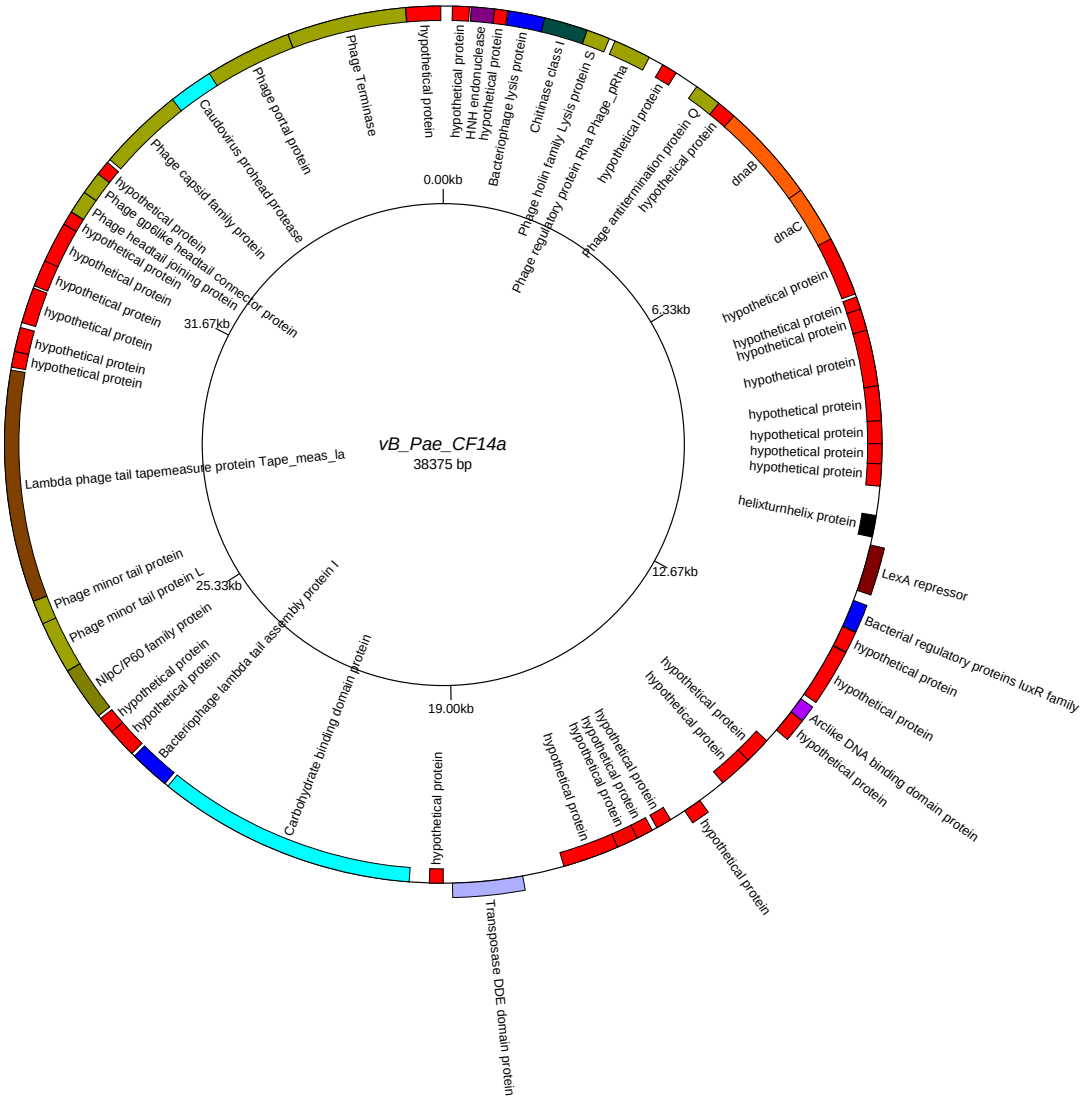


Figure 14: Genome map of vB_Pae_CF14a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The potential incomplete temperate phage vB_Pae_CF14b genome size was 28 kb, the phage was assembled using SPAdes and has a mean coverage of 27. The genome map of this phage is shown in Figure 15. The total number of genes identified was 45 of which 7 were identified with putative functions these can be seen in Table 15. Blastn on viruses only database showed a 2.5 kb / 2.6 kb match with *Pseudomonas* phage B3 (GenBank: NC_006548.1). The phage vB_Pae_CF14b had a 63.3 % G-C content. No tRNA gene was identified however; several genes associated with regulatory proteins were identified and an Integrase gene was identified.

Start	End	Direction	Putative functional protein
470	111	Reverse	Mor transcription activator family protein
7989	6205	Reverse	Integrase core domain protein
10998	11387	Forward	Helix-turn-helix domain protein
13259	13888	Forward	Soluble lytic murein transglycosylase precursor
15954	17657	Forward	Terminase-like family protein
19135	20385	Forward	Phage Mu protein F like protein
20382	20954	Forward	Phage virion morphogenesis family protein

Table 15: Functional genes identified for vB_Pae_CF14b.

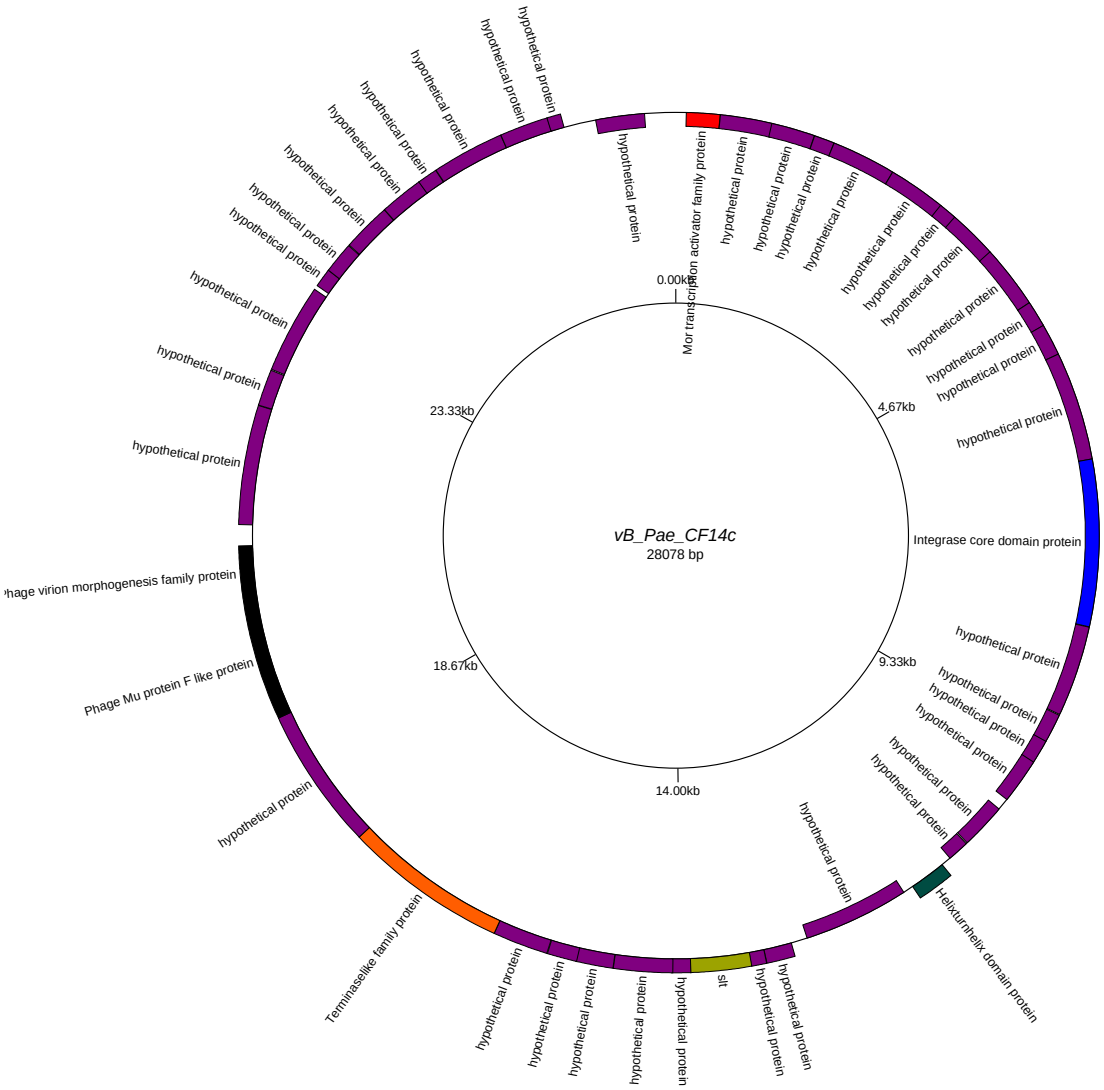


Figure 15: Genome map of vB_Pae_CF14b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The potential incomplete temperate phage vB_Pae_CF14c genome size was 26.1 kb, the phage was assembled using SPAdes and has a mean coverage of 151. The genome map of this phage is shown in Figure 16. The total number of genes identified was 45 of which 7 were identified with putative functions these can be seen in Table 16. Blastn on viruses only database showed a 17 kb / 18.2 kb match with *Pseudomonas* phage JBD24 (GenBank: NC_020203.1). The phage vB_Pae_CF14c had a 64.5 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
5042	4134	Reverse	Mu-like prophage major head subunit gpT
6584	5478	Reverse	Mu-like prophage I protein
8070	7603	Reverse	Phage virion morphogenesis family protein
9356	8070	Reverse	Phage Mu protein F like protein
16274	15825	Reverse	Mor transcription activator family protein
19596	19078	Reverse	Bacteriophage Mu Gam like protein
23803	21809	Reverse	Mu DNA-binding domain protein
24861	24505	Reverse	DNA-binding transcriptional regulator Nlp
25031	25744	Forward	HTH-type transcriptional regulator PrtR

Table 16: Functional genes identified for vB_Pae_CF14c.

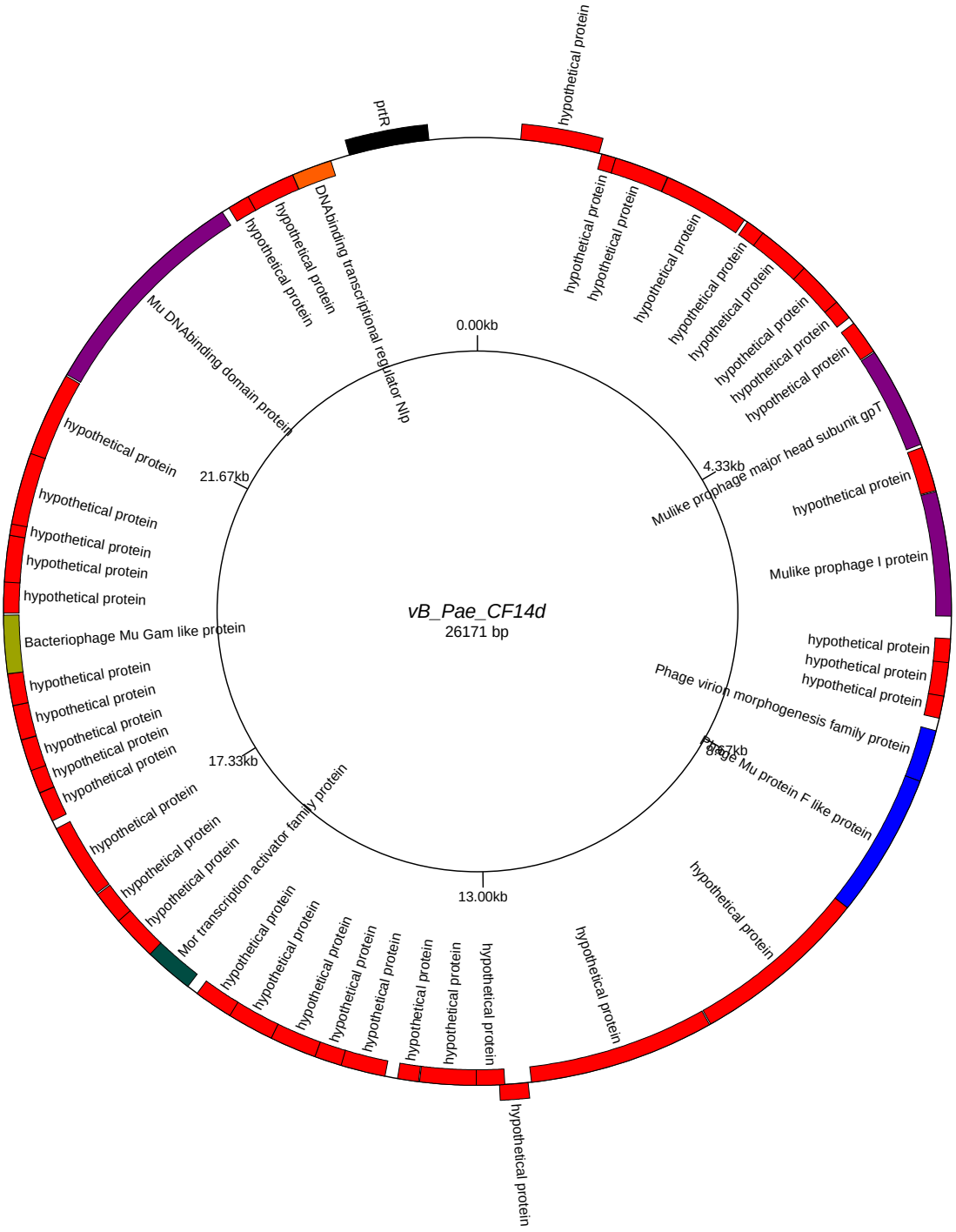


Figure 16: Genome map of vB_Pae_CF14c produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The potential incomplete temperate phage vB_Pae_CF15a genome size was 41.1 kb, the phage was assembled using SPAdes and has a mean coverage of 43. The genome map of this phage is shown in Figure 17. The total number of genes identified was 63 of which 27 were identified with putative functions these can be seen in Table 17. Blastn on viruses only database showed a 1.9 kb / 1.9 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF15a had a 61.2 % G-C content. No tRNA gene was identified however; genes associated with structural, lysogenic and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
874	1194	Forward	Cro
5462	6277	Forward	DNA replication protein DnaC
6274	7671	Forward	Replicative DNA helicase
7943	8332	Forward	Phage antitermination protein Q
9107	9637	Forward	Phage regulatory protein Rha (Phage_pRha)
9722	10054	Forward	Phage holin family (Lysis protein S)
10051	10668	Forward	Chitinase class I
10668	11180	Forward	Bacteriophage lysis protein
11375	11704	Forward	HNH endonuclease
12613	14292	Forward	Phage Terminase
14295	15518	Forward	Phage portal protein
15502	16146	Forward	Caudovirus prohead protease
16143	17357	Forward	Phage capsid family protein
17612	17935	Forward	Phage gp6-like head-tail connector protein
17935	18261	Forward	Phage head-tail joining protein
20633	23908	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
23908	24246	Forward	Phage minor tail protein
24243	24989	Forward	Phage minor tail protein L
24992	25750	Forward	NlpC/P60 family protein
26501	27073	Forward	Bacteriophage lambda tail assembly protein I
27130	30795	Forward	Carbohydrate binding domain protein

32383	31400	Reverse	Transposase DDE domain protein
37091	36861	Reverse	Arc-like DNA binding domain protein
37951	37166	Reverse	BRO family, N-terminal domain
38649	38263	Reverse	Bacterial regulatory proteins, luxR family
39457	38798	Reverse	LexA repressor
39567	39872	Forward	Helix-turn-helix

Table 17: Functional genes identified for vB_Pae_CF15a.

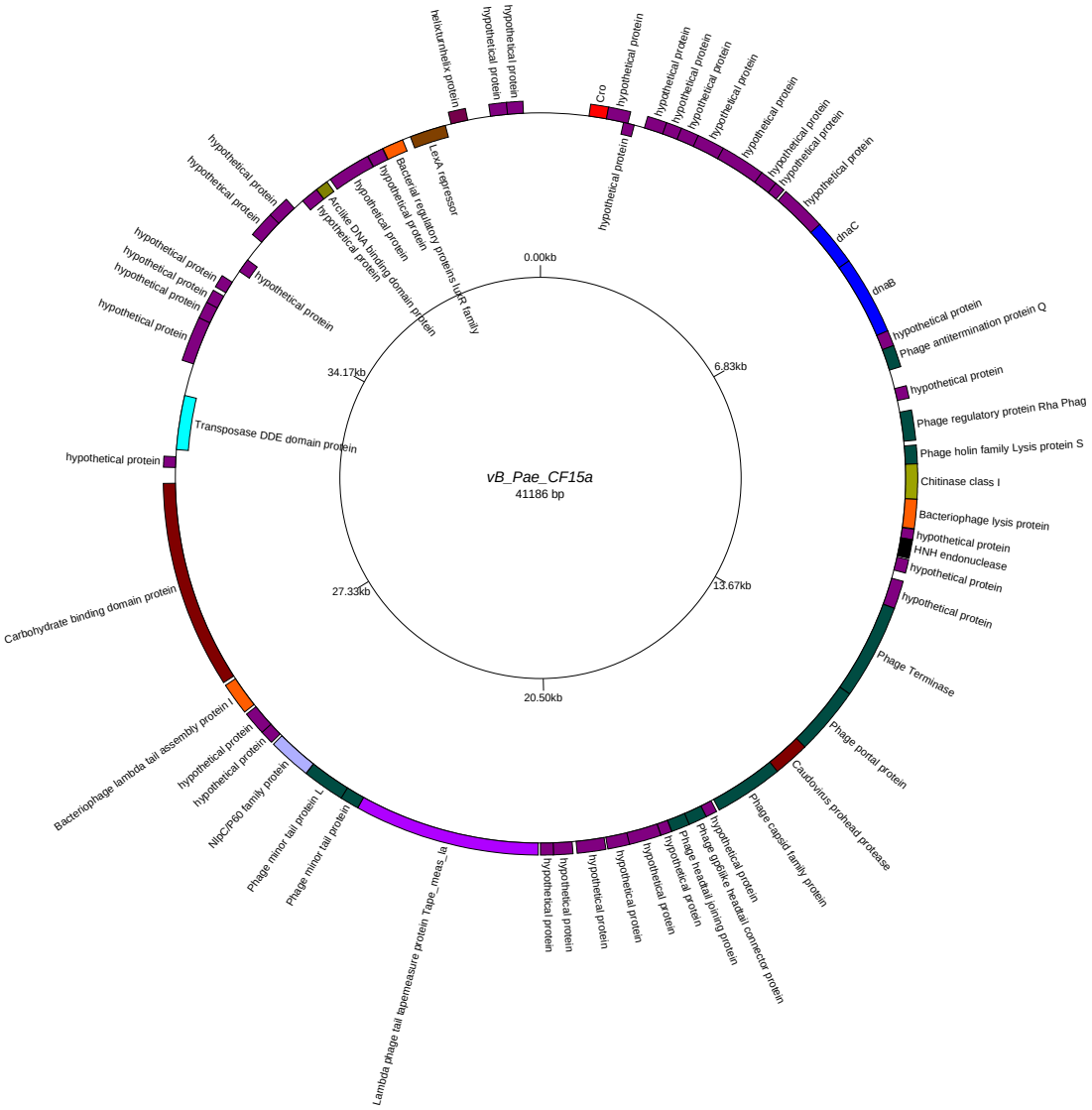


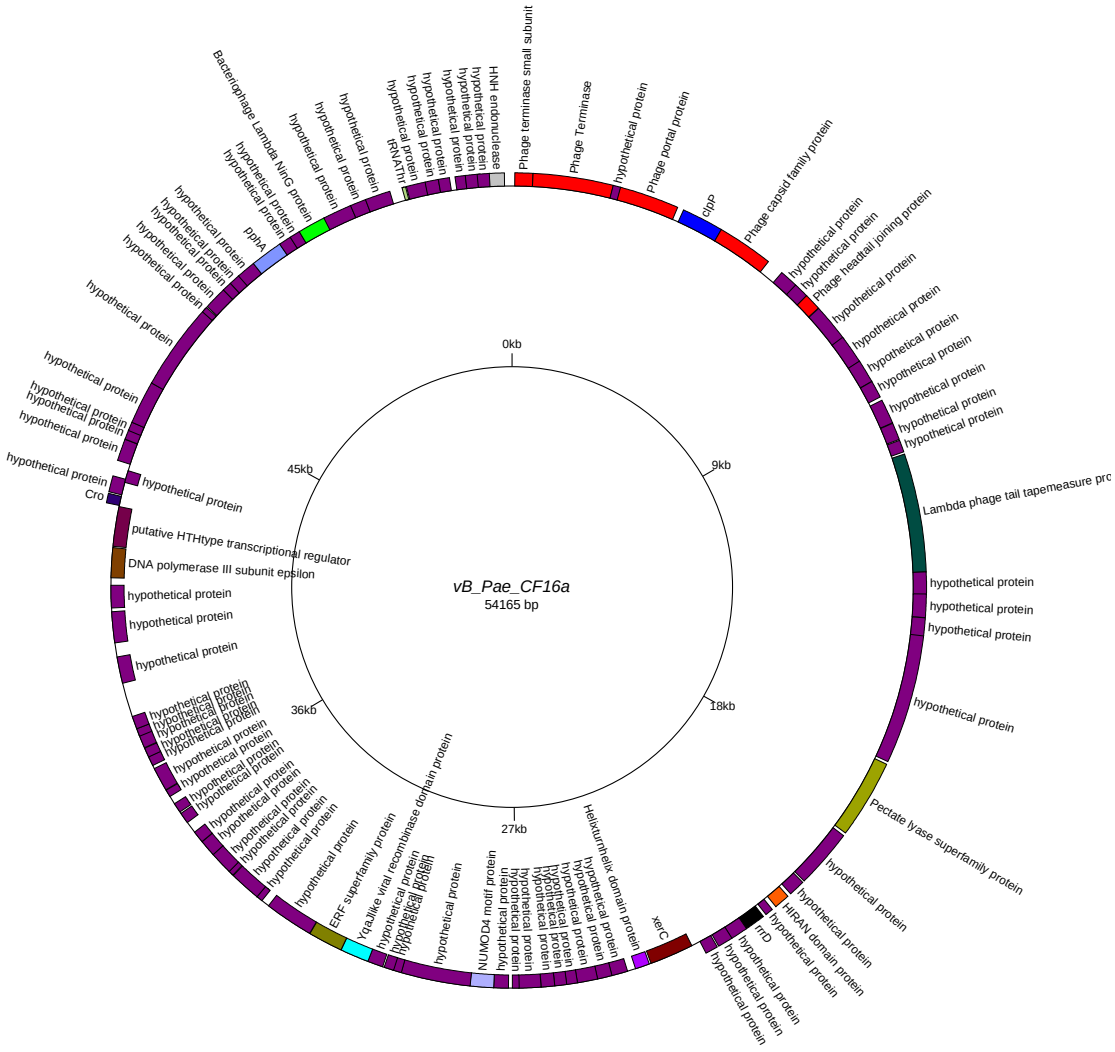
Figure 17: Genome map of vB_Pae_CF15a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF16a genome size was 54.1 kb, the phage was assembled using SPAdes and has a mean coverage of 39. The genome map of this phage is shown in Figure 18. The total number of genes identified was 97 of which 22 were identified with putative functions these can be seen in Table 18. Blastn on viruses only database showed a 6.3 kb / 7 kb match with *Pseudomonas* phage vB_PaeS_PMG1 (GenBank: NC_016765.1). The phage vB_Pae_CF16a had a 58.4 % G-C content. A tRNA gene was identified along with genes associated with structural, lysogenic and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
56	436	Forward	Phage terminase, small subunit
438	2129	Forward	Phage Terminase
2283	3545	Forward	Phage portal protein
3677	4567	Forward	ATP-dependent Clp protease proteolytic subunit
4564	5751	Forward	Phage capsid family protein
6816	7172	Forward	Phage head-tail joining protein
10722	13226	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
17351	19021	Forward	Pectate lyase superfamily protein
20804	21136	Forward	HIRAN domain protein
21467	21901	Forward	Lysozyme RrrD
24034	23078	Reverse	Tyrosine recombinase XerC
24387	24073	Reverse	Helix-turn-helix domain protein
27993	27478	Reverse	NUMOD4 motif protein
30877	30251	Reverse	YqaJ-like viral recombinase domain protein
31633	30881	Reverse	ERF superfamily protein
41473	40823	Reverse	DNA polymerase III subunit epsilon
42366	41503	Reverse	putative HTH-type transcriptional regulator
42391	42591	Forward	Cro
48351	49037	Forward	Serine/threonine-protein phosphatase 1
49515	50114	Forward	Bacteriophage Lambda NinG protein
51847	51922	Forward	tRNA-Thr(tgt)

53691	54005	Forward	HNH endonuclease

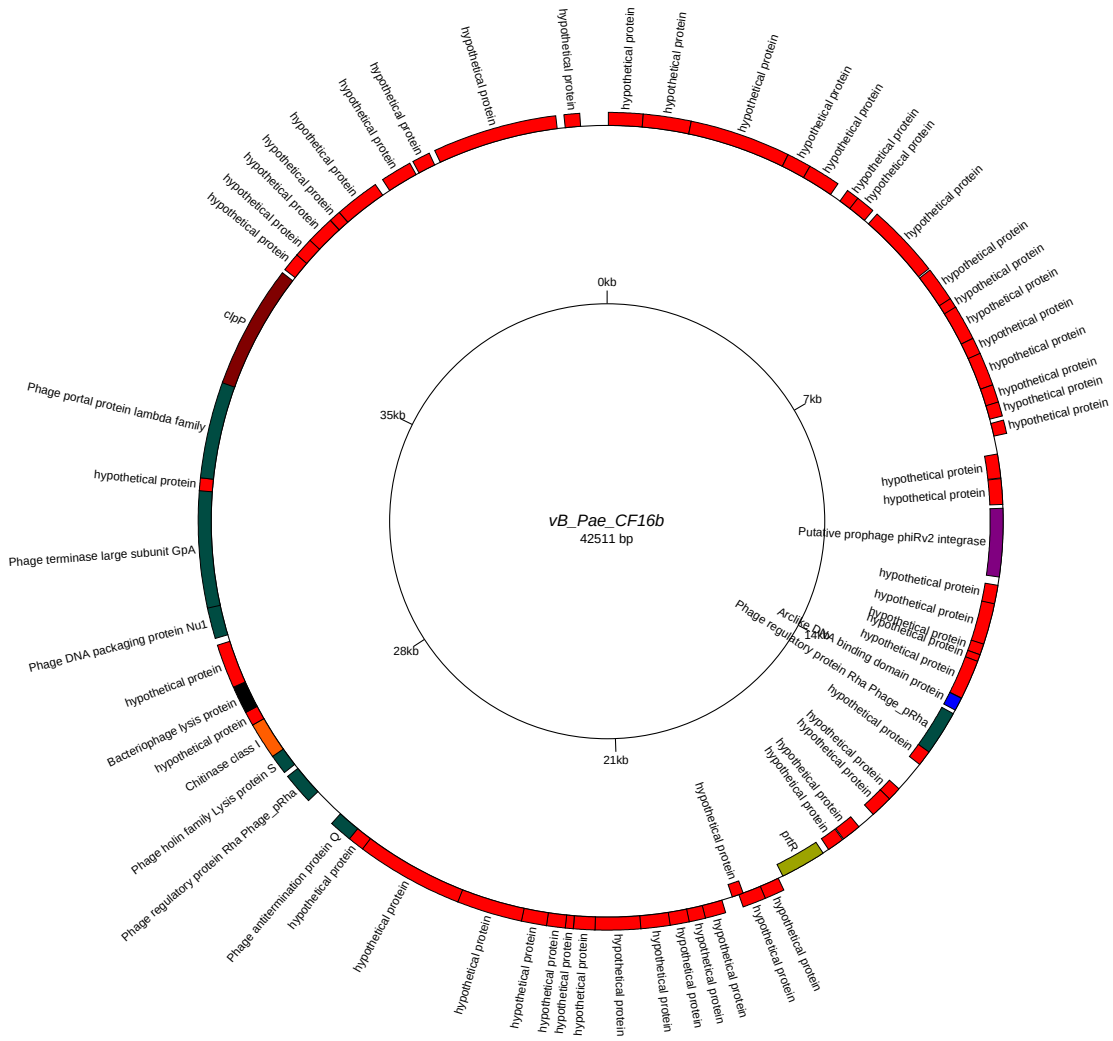
Table 18: Functional genes identified for vB_Pae_CF16a.


vB_Pae_CF16a
54165 bp
0kb
9kb
18kb
27kb
36kb
45kb
Phage terminase small subunit
Phage Terminase
hypothetical protein
Phage portal protein
clpP
Phage capsid family protein
hypothetical protein
hypothetical protein
Phage headtail joining protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Lambda phage tail tapemeasure prc
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Pectate lyase superfamily protein
hypothetical protein
HIRAN domain protein
hypothetical protein
rnlD
hypothetical protein
hypothetical protein
xerC
Helixturnhelix domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Pae_CF16b genome size was 42.5 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 19. The total number of genes identified was 68 of which 13 were identified with putative functions these can be seen in Table 19. Blastn on viruses only database showed an 8 kb / 8.6 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF16b had a 61.5 % G-C content. No tRNA gene was identified however; genes associated with structural, lysogenic and regulatory proteins were identified, including an integrase putative protein.

Start	End	Direction	Putative functional protein
11583	10405	Reverse	Putative prophage phiRv2 integrase
13993	13763	Reverse	Arc-like DNA binding domain protein
14841	14062	Reverse	Phage regulatory protein Rha (Phage_pRha)
18145	17354	Reverse	HTH-type transcriptional regulator PrtR
25868	26257	Forward	Phage antitermination protein Q
26796	27323	Forward	Phage regulatory protein Rha (Phage_pRha)
27408	27740	Forward	Phage holin family (Lysis protein S)
27737	28354	Forward	Chitinase class I
28590	29060	Forward	Bacteriophage lysis protein
29917	30462	Forward	Phage DNA packaging protein Nu1
30434	32398	Forward	Phage terminase large subunit (GpA)
32604	34250	Forward	Phage portal protein, lambda family
34222	36303	Forward	ATP-dependent Clp protease proteolytic subunit

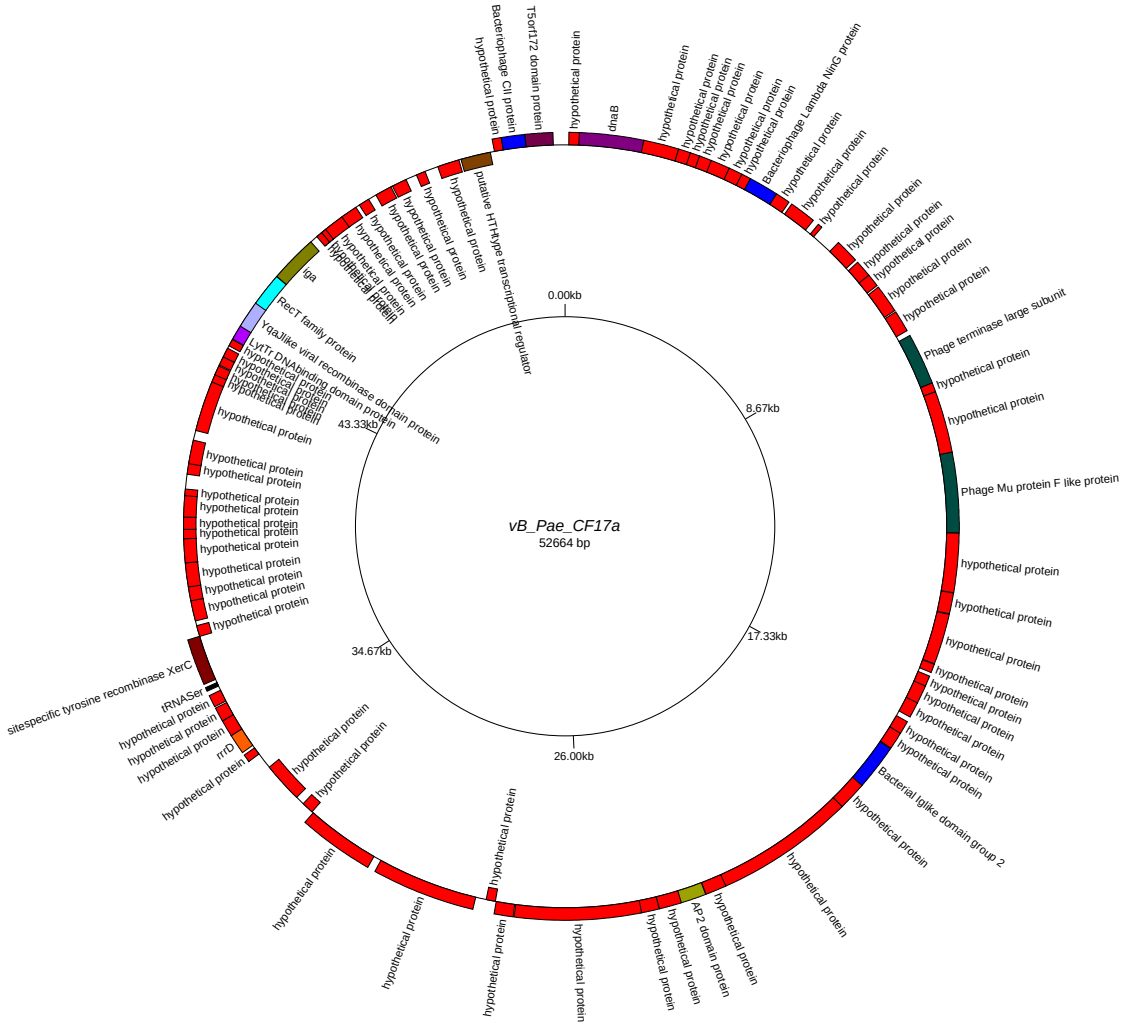
Table 19: Functional genes identified for vB_Pae_CF16b.


vB_Pae_CF16b
42511 bp
0kb
7kb
14kb
21kb
28kb
35kb
clpP
Phage portal protein lambda family
hypothetical protein
Phage terminase large subunit GpA
Phage DNA packaging protein Nu1
hypothetical protein
Bacteriophage lysis protein
hypothetical protein
Chitinase class I
Phage holin family Lysis protein S
Phage regulatory protein Rha Phage_pRha
Phage antitermination protein Q
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Pae_CF17a genome size was 52.6 kb, the phage was assembled using SPAdes and has a mean coverage of 208. The genome map of this phage is shown in Figure 20. The total number of genes identified was 86 of which 18 were identified with putative functions these can be seen in Table 20. Blastn on viruses only database showed an 8 kb / 8.6 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_CF17a had a 58.9 % G-C content. A tRNA gene was identified along with genes associated with structural, lysogenic and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
303	1688	Forward	Replicative DNA helicase
4102	4749	Forward	Bacteriophage Lambda NinG protein
8927	10045	Forward	Phage terminase large subunit
11578	13308	Forward	Phage Mu protein F like protein
18193	19188	Forward	Bacterial Ig-like domain (group 2)
23274	23834	Forward	AP2 domain protein
30729	32375	Forward	D-glucuronyl C5-epimerase C-terminus
33769	32870	Reverse	BRO family, N-terminal domain
34372	34806	Forward	Lysozyme RrrD
35876	35963	Forward	tRNA-Ser(cga)
36037	37047	Forward	site-specific tyrosine recombinase XerC
44134	43799	Reverse	LytTr DNA-binding domain protein
44751	44131	Reverse	YqaJ-like viral recombinase domain protein
45481	44735	Reverse	RecT family protein
46553	45489	Reverse	IgA-specific serine endopeptidase autotransporter precursor
51012	50347	Reverse	putative HTH-type transcriptional regulator
51302	51805	Forward	Bacteriophage CII protein
51809	52405	Forward	T5orf172 domain protein

Table 20: Functional genes identified for vB_Pae_CF17a.


vB_Pae_CF17a
52664 bp
0.00kb
8.67kb
17.33kb
26.00kb
34.67kb
43.33kb
Bacteriophage CII protein
T5orf172 domain protein
hypothetical protein
hypothetical protein
dnaB
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Bacteriophage Lambda NinG protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage terminase large subunit
hypothetical protein
hypothetical protein
Phage Mu protein F like protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Bacterial IgIike domain group 2
hypothetical protein
hypothetical protein
hypothetical protein
AP2 domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Pae_CF17b genome size was 37.7 kb, the phage was assembled using SPAdes and has a mean coverage of 379. The genome map of this phage is shown in Figure 21. The total number of genes identified was 58 of which 18 were identified with putative functions these can be seen in Table 21. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF17b had a 64.3 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
1203	490	Reverse	HTH-type transcriptional regulator PrtR
1373	1729	Forward	DNA-binding transcriptional regulator Nlp
2431	4425	Forward	Mu DNA-binding domain protein
6638	7156	Forward	Bacteriophage Mu Gam like protein
9960	10409	Forward	Mor transcription activator family protein
16878	18164	Forward	Phage Mu protein F like protein
18164	18631	Forward	Phage virion morphogenesis family protein
19650	20756	Forward	Mu-like prophage I protein
21192	22100	Forward	Mu-like prophage major head subunit gpT
25132	28692	Forward	Prophage tail length tape measure protein

Table 21: Functional genes identified for vB_Pae_CF17b.

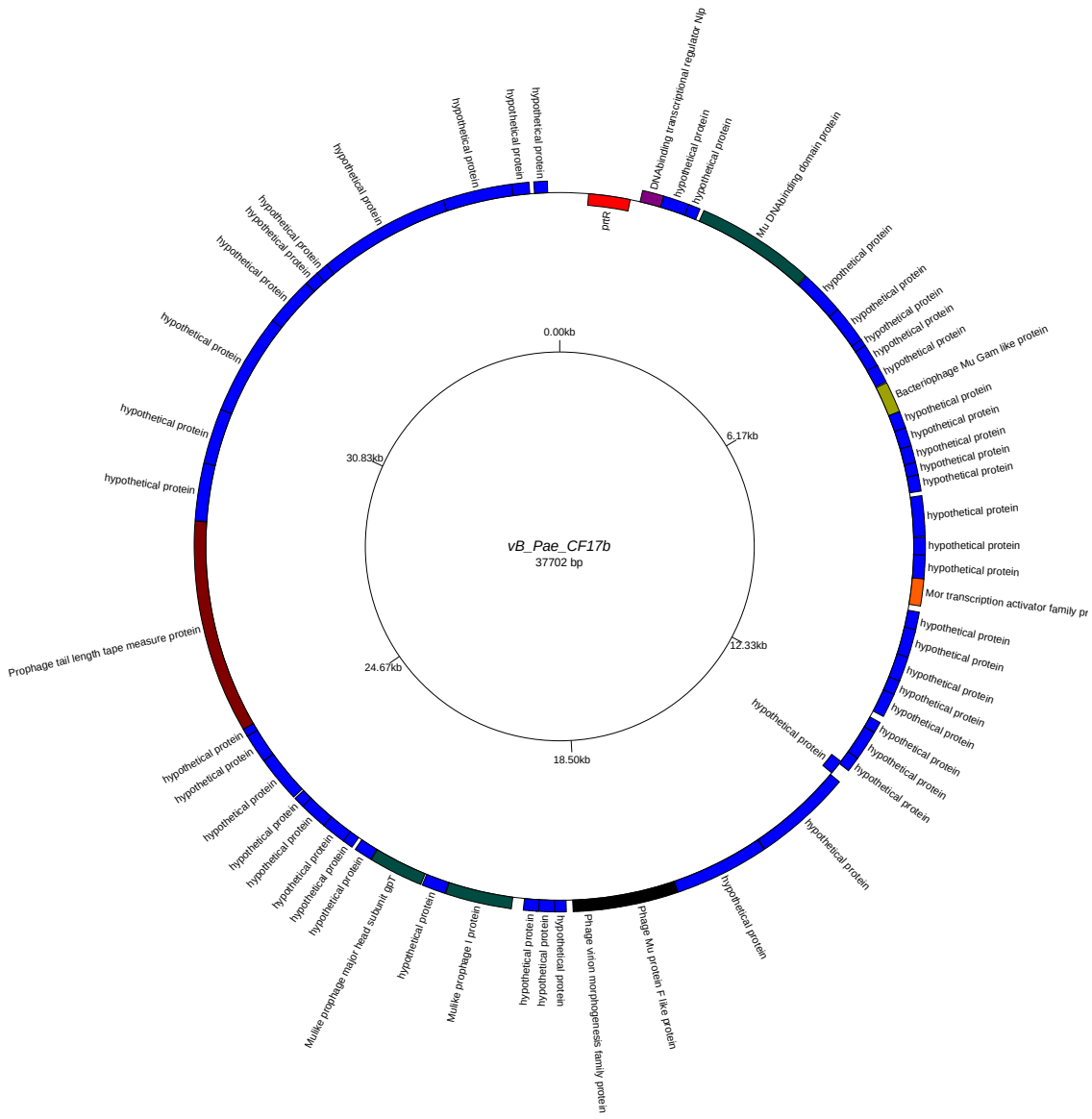


Figure 21: Genome map of vB_Pae_CF17b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The potentially incomplete temperate phage vB_Pae_CF18a genome size was 35.2 kb, the phage was assembled using SPAdes and has a mean coverage of 1549. The genome map of this phage is shown in Figure 22. The total number of genes identified was 59 of which 14 were identified with putative functions these can be seen in Table 22. Blastn on viruses only database showed a 6.2 kb / 6.7 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF18a had a 61.6 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified. An integrase gene was also predicted.

Start	End	Direction	Putative functional protein
9170	7992	Reverse	Putative prophage phiRv2 integrase
11580	11350	Reverse	Arc-like DNA binding domain protein
12427	11660	Reverse	Phage regulatory protein Rha (Phage_pRha)
13832	13446	Reverse	Bacterial regulatory proteins, luxR family
15706	14915	Reverse	HTH-type transcriptional regulator PrtR
23429	23818	Forward	Phage antitermination protein Q
24357	24884	Forward	Phage regulatory protein Rha (Phage_pRha)
24969	25301	Forward	Phage holin family (Lysis protein S)
25298	25915	Forward	Chitinase class I
26151	26621	Forward	Bacteriophage lysis protein
27478	28023	Forward	Phage DNA packaging protein Nu1
27995	29959	Forward	Phage terminase large subunit (GpA)
30165	31811	Forward	Phage portal protein, lambda family
31783	33864	Forward	ATP-dependent Clp protease proteolytic subunit

Table 22: Functional genes identified for vB_Pae_CF18a.

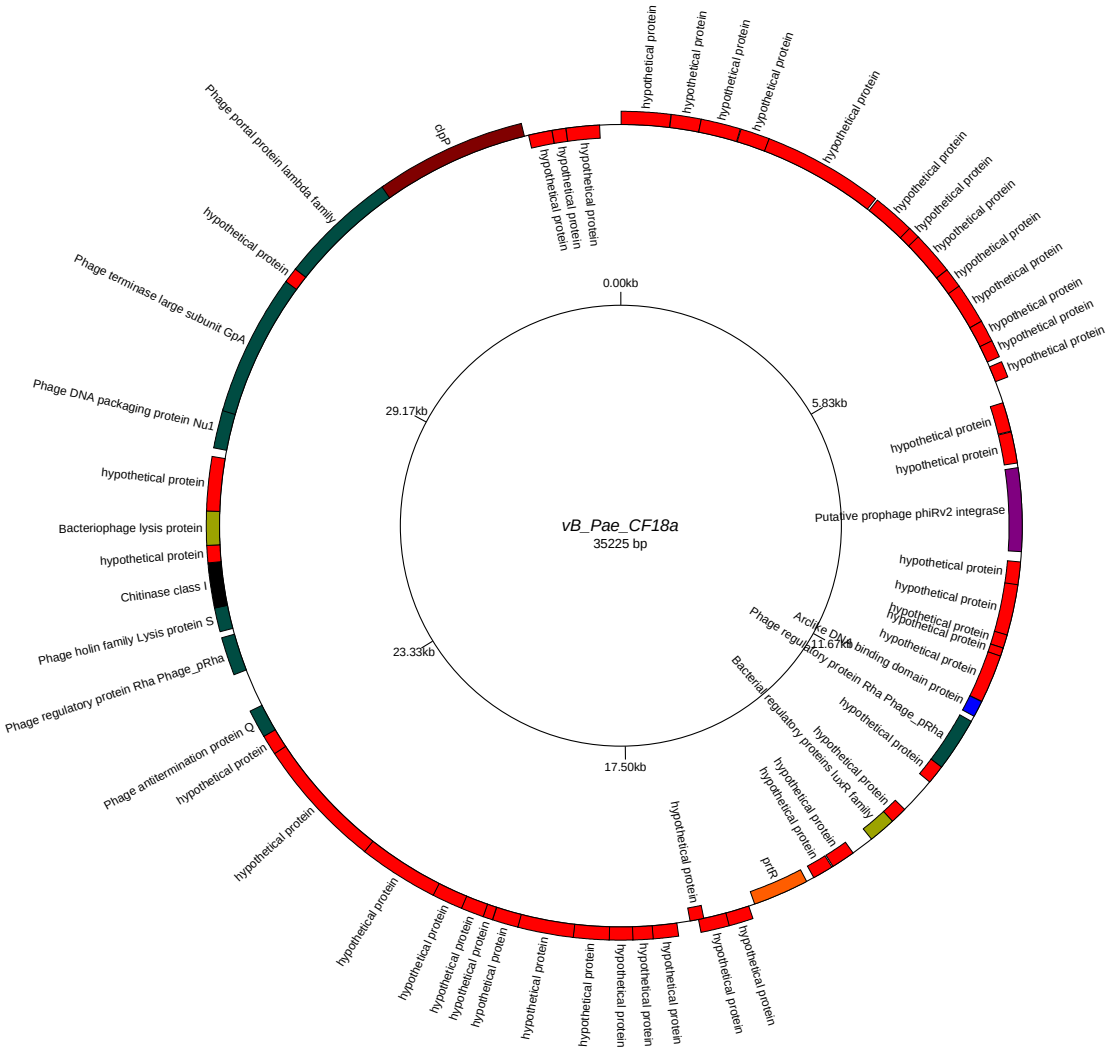


Figure 22: Genome map of vB_Pae_CF18a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF19a genome size was 62.6 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 23. The total number of genes identified was 63 of which 14 were identified with putative functions these can be seen in Table 23. Blastn on viruses only database showed a 13 kb / 13.2 kb match with *Pseudomonas* phage LKA5 (GenBank: KC900378.1). The phage vB_Pae_CF19a had a 63.7 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified. An integrase gene was also predicted.

Start	End	Direction	Putative functional protein
25515	26054	Forward	Phage lysozyme
29120	30355	Forward	Putative prophage CPS-53 integrase
30589	30356	Reverse	Response regulator inhibitor for tor operon
34926	33118	Reverse	C-5 cytosine-specific DNA methylase
37553	34923	Reverse	DNA methylase
37907	37698	Reverse	LexA repressor
39124	37904	Reverse	Recombination-associated protein RdgC
39648	39214	Reverse	Single-stranded DNA-binding protein
42016	40379	Reverse	YqaJ-like viral recombinase domain protein
42711	42013	Reverse	ERF superfamily protein
44808	44437	Reverse	Carbon storage regulator homolog
49438	50052	Forward	Bacteriophage Lambda NinG protein
51568	52851	Forward	Phage terminase large subunit
58364	59236	Forward	RyR domain protein

Table 23: Functional genes identified for vB_Pae_CF19a.

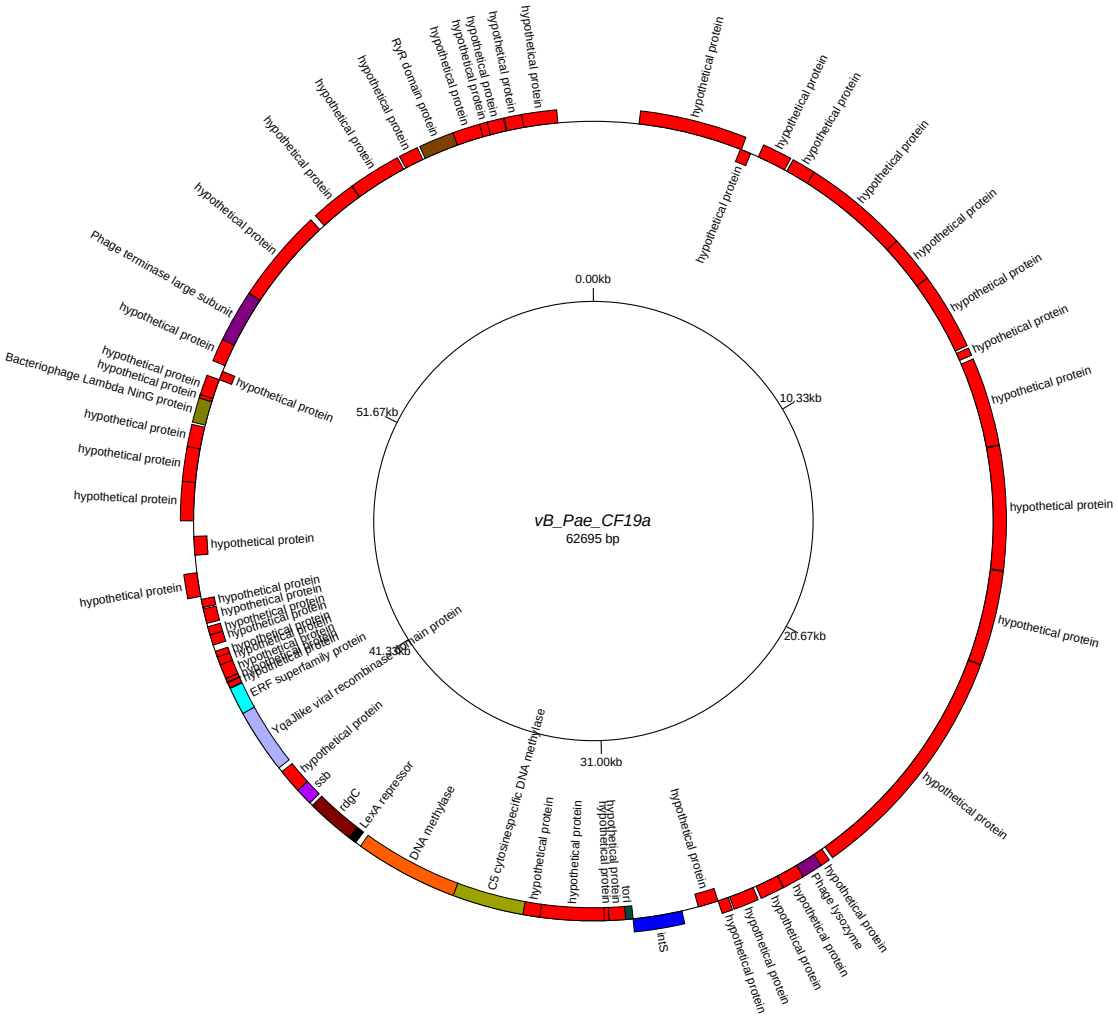


Figure 23: Genome map of vB_Pae_CF19a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF20a genome size was 52.6 kb, the phage was assembled using SPAdes with a mean coverage of 195. The genome map of this phage is shown in Figure 24. The total number of genes identified was 86 of which 18 were identified with putative functions these can be seen in Table 24. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_CF20a had a 58.9 % G-C content. A tRNA gene was identified along with genes associated with structural, lysogenic and regulatory proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
867	271	Reverse	T5orf172 domain protein
1374	871	Reverse	Bacteriophage CII protein
1664	2329	Forward	putative HTH-type transcriptional regulator
6123	7187	Forward	IgA-specific serine endopeptidase autotransporter precursor
7195	7941	Forward	RecT family protein
7925	8545	Forward	YqaJ-like viral recombinase domain protein
8542	8877	Forward	LytTr DNA-binding domain protein
16639	15629	Reverse	site-specific tyrosine recombinase XerC
16800	16713	Reverse	tRNA-Ser(cga)
18304	17870	Reverse	Lysozyme RrrD
18907	19806	Forward	BRO family, N-terminal domain
21947	20301	Reverse	D-glucuronyl C5-epimerase C-terminus
29402	28842	Reverse	AP2 domain protein
34483	33488	Reverse	Bacterial Ig-like domain (group 2)
41098	39368	Reverse	Phage Mu protein F like protein
43749	42631	Reverse	Phage terminase large subunit
48574	47927	Reverse	Bacteriophage Lambda NinG protein
52373	50988	Reverse	Replicative DNA helicase

Table 24: Functional genes identified for vB_Pae_CF20a.

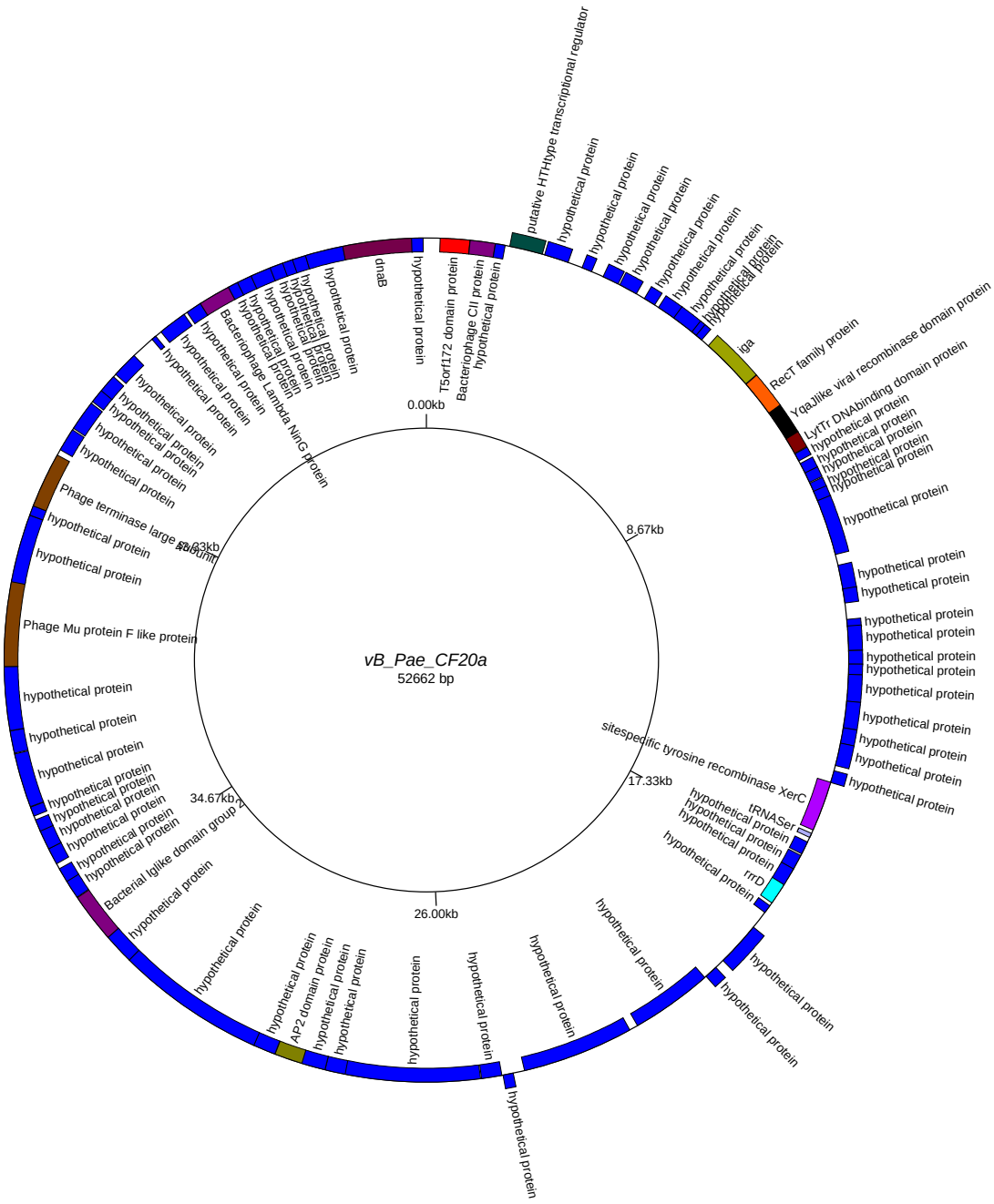
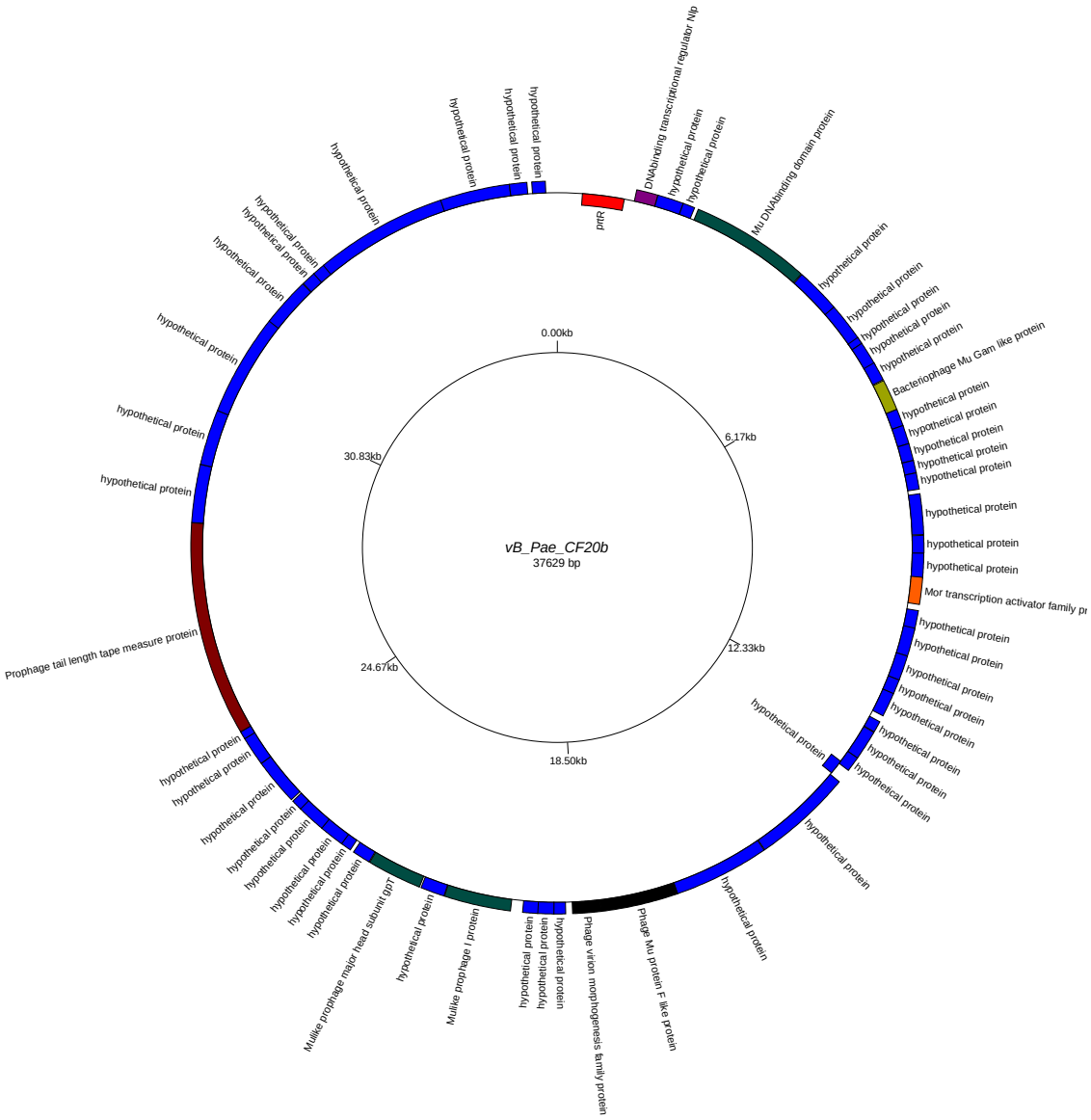


Figure 24: Genome map of vB_Pae_CF20a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF20b genome size was 37.6 kb, the phage was assembled using SPAdes and has a mean coverage of 144. The genome map of this phage is shown in Figure 25. The total number of genes identified was 58 of which 18 were identified with putative functions these can be seen in Table 25. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF20b had a 64.3 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
1137	424	Reverse	HTH-type transcriptional regulator PrtR
1307	1663	Forward	DNA-binding transcriptional regulator Nlp
2365	4359	Forward	Mu DNA-binding domain protein
6572	7090	Forward	Bacteriophage Mu Gam like protein
9894	10343	Forward	Mor transcription activator family protein
16812	18098	Forward	Phage Mu protein F like protein
18098	18565	Forward	Phage virion morphogenesis family protein
19584	20690	Forward	Mu-like prophage I protein
21126	22034	Forward	Mu-like prophage major head subunit gpT
25066	28626	Forward	Prophage tail length tape measure protein

Table 25: Functional genes identified for vB_Pae_CF20b.


vB_Pae_CF20b
37629 bp
0.00kb
6.17kb
12.33kb
18.50kb
24.67kb
30.83kb
DNAbinding transcriptional regulator Nip
hypothetical protein
hypothetical protein
Mu DNAbinding domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Bacteriophage Mu Gam like protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Mor transcription activator family pr
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage Mu protein F like protein
Phage virion morphogenesis family protein
hypothetical protein
hypothetical protein
Mulike prophage I protein
hypothetical protein
Mulike prophage major head subunit gpT
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Prophage tail length tape measure protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The potentially incomplete temperate phage vB_Pae_CF21a genome size was 35.4 kb, the phage was assembled using SPAdes and extended using PriceTI. The genome map of this phage is shown in Figure 26. The total number of genes identified was 56 of which 14 were identified with putative functions these can be seen in Table 26. Blastn on viruses only database showed a 9.6 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF21a had a 61.3 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified. An integrase gene was also predicted.

Start	End	Direction	Putative functional protein
5885	6676	Forward	HTH-type transcriptional regulator PrtR
7759	8145	Forward	Bacterial regulatory proteins, luxR family
9164	9931	Forward	Phage regulatory protein Rha (Phage_pRha)
10011	10241	Forward	Arc-like DNA binding domain protein
12421	13599	Forward	Putative prophage phiRv2 integrase
31000	28919	Reverse	ATP-dependent Clp protease proteolytic subunit
32618	30972	Reverse	Phage portal protein, lambda family
34788	32824	Reverse	Phage terminase large subunit (GpA)
35392	34760	Reverse	Phage DNA packaging protein Nu1

Table 26: Functional genes identified for vB_Pae_CF21a.

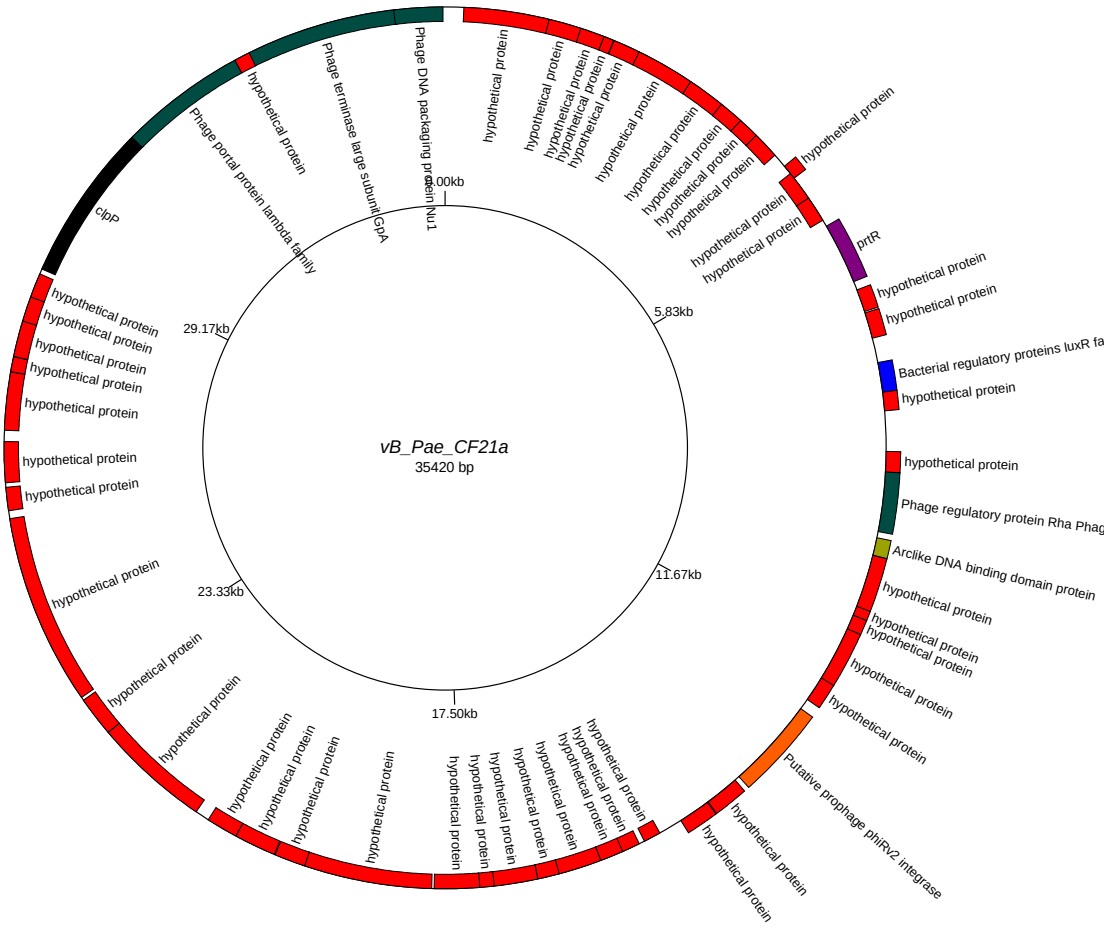


Figure 26: Genome map of vB_Pae_CF21a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF22a genome size was 61.7 kb, the phage was assembled using SPAdes and has a mean coverage of 16. The genome map of this phage is shown in Figure 27. The total number of genes identified was 62 of which 14 were identified with putative functions these can be seen in Table 27. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66 (GenBank: KC262634.1). The phage vB_Pae_CF22a had a 63.7 % G-C content. No tRNA gene was identified however; a CPS-53 integrase like gene was identified.

Start	End	Direction	Putative functional protein
4605	5102	Forward	Single-stranded DNA-binding protein
5129	6349	Forward	Recombination-associated protein RdgC
6346	6555	Forward	LexA repressor
6700	9330	Forward	DNA methylase
9327	11135	Forward	C-5 cytosine-specific DNA methylase
13664	13897	Forward	Response regulator inhibitor for tor operon
15133	13898	Reverse	Putative prophage CPS-53 integrase
18738	18199	Reverse	Phage lysozyme
47162	46290	Reverse	RyR domain protein
53961	52678	Reverse	Phage terminase large subunit
55759	55145	Reverse	Bacteriophage Lambda NinG protein
58365	58165	Reverse	Cro
58473	59270	Forward	HTH-type transcriptional regulator PrtR
60689	61060	Forward	Carbon storage regulator homolog

Table 27: Functional genes identified for vB_Pae_CF22a.

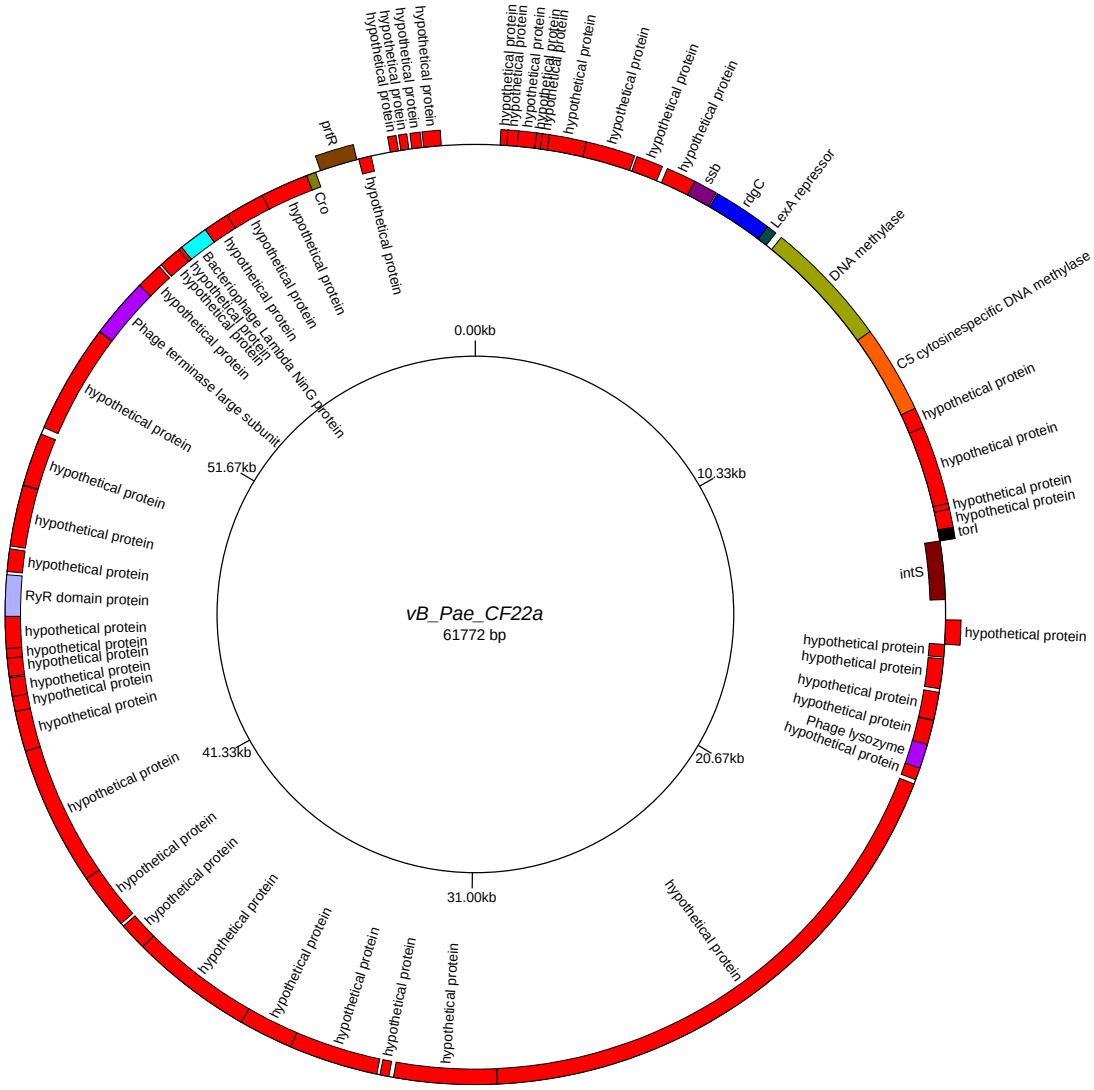
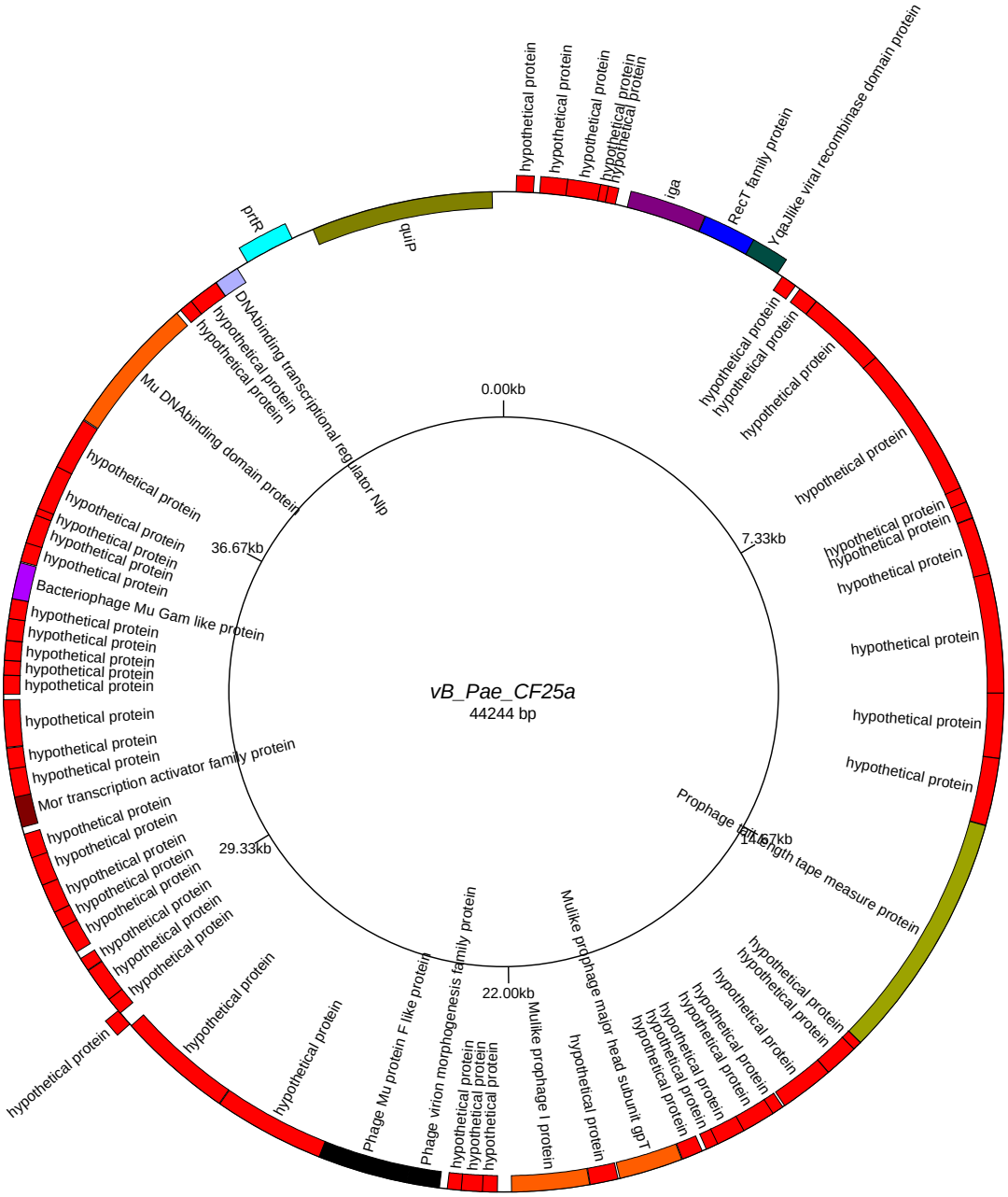


Figure 27: Genome map of vB_Pae_CF22a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF25a genome size was 44.2 kb, the phage was assembled using SPAdes and has a mean coverage of 103. The genome map of this phage is shown in Figure 28. The total number of genes identified was 67 of which 14 were identified with putative functions these can be seen in Table 28. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF25a had a 64.1 % G-C content. No tRNA gene was identified however; Mu-like phage structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
1757	2821	Forward	IgA-specific serine endopeptidase autotransporter precursor
2829	3575	Forward	RecT family protein
3559	4080	Forward	YqaJ-like viral recombinase domain protein
16529	12969	Reverse	Prophage tail length tape measure protein
20469	19561	Reverse	Mu-like prophage major head subunit gpT
22011	20905	Reverse	Mu-like prophage I protein
23497	23030	Reverse	Phage virion morphogenesis family protein
24783	23497	Reverse	Phage Mu protein F like protein
31701	31252	Reverse	Mor transcription activator family protein
35023	34505	Reverse	Bacteriophage Mu Gam like protein
39230	37236	Reverse	Mu DNA-binding domain protein
40288	39932	Reverse	DNA-binding transcriptional regulator Nlp
40458	41171	Forward	HTH-type transcriptional regulator PrtR
44082	41491	Reverse	Acyl-homoserine lactone acylase QuiP precursor

Table 28: Functional genes identified for vB_Pae_CF25a.


vB_Pae_CF25a
44244 bp
0.00kb
7.33kb
14.67kb
22.00kb
29.33kb
36.67kb
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
iga
RecT family protein
YqaJlike viral recombinase domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The potential partial temperate phage vB_Pae_CF25b genome size was 34.6 kb, the phage was assembled using SPAdes and extended using PriceTI. The genome map of this phage is shown in Figure 29. The total number of genes identified was 46 of which 8 were identified with putative functions these can be seen in Table 29. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_CF25b had a 59.1 % G-C content. No tRNA gene was identified however; an Ig-like domain was identified.

Start	**End**	**Direction**	**Putative functional protein**
405	1790	Forward	Replicative DNA helicase
4204	4851	Forward	Bacteriophage Lambda NinG protein
9029	10147	Forward	Phage terminase large subunit
11680	13410	Forward	Phage Mu protein F like protein
18295	19290	Forward	Bacterial Ig-like domain (group 2)
23376	23936	Forward	AP2 domain protein
30831	32477	Forward	D-glucuronyl C5-epimerase C-terminus
33832	32972	Reverse	BRO family, N-terminal domain

Table 29: Functional genes identified for vB_Pae_CF25b.

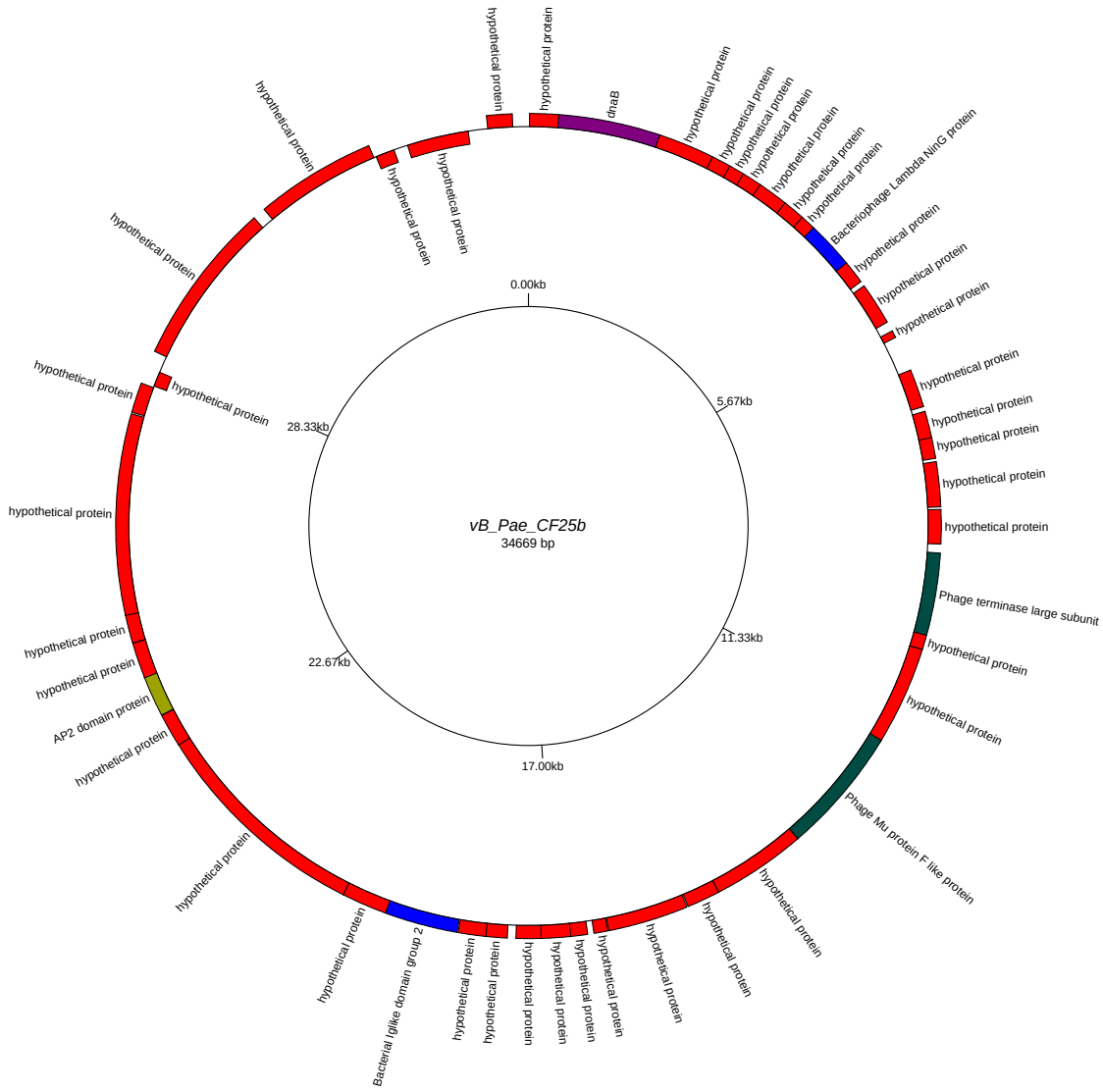
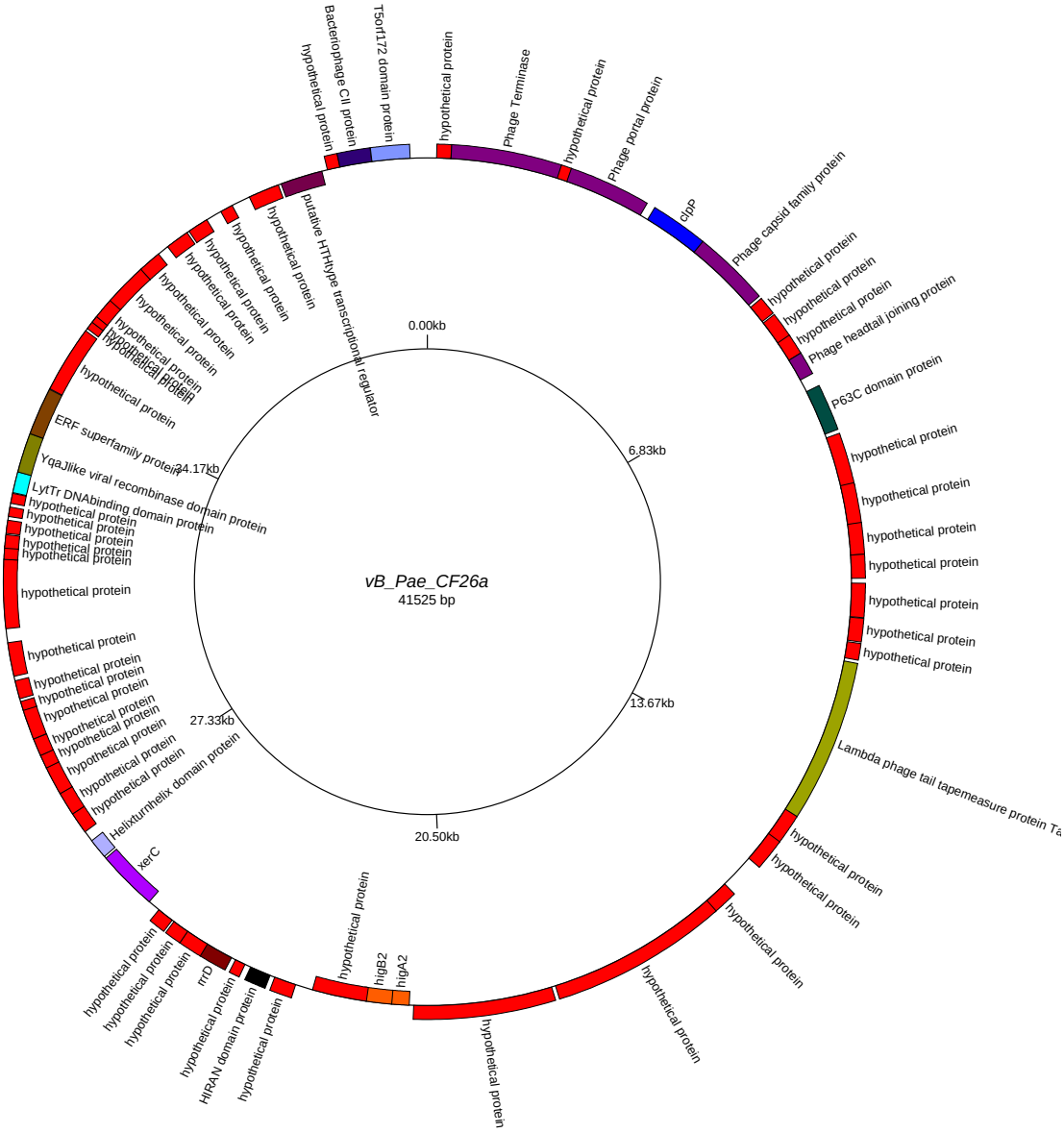


Figure 29: Genome map of vB_Pae_CF25b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF26a genome size was 41.5 kb, the phage was assembled using SPAdes all the reads were used and gave a mean coverage of 52. The genome map of the phage is shown in Figure 30. The total number of genes identified was 68 of which 20 were identified with putative functions shown in Table 30. Blastn on viruses only database showed a 4.9 kb / 5.1 kb match with *Pseudomonas* phage vB_PaeS_PMG1 (GenBank: NC_016765.1). The phage vB_Pae_CF26a had a 58.3 % G-C content. No tRNA gene was identified however; structural and lysogenic genes were identified.

Start	End	Direction	Putative functional protein
369	2060	Forward	Phage Terminase
2214	3476	Forward	Phage portal protein
3608	4498	Forward	ATP-dependent Clp protease proteolytic subunit
4495	5682	Forward	Phage capsid family protein
6747	7103	Forward	Phage head-tail joining protein
7308	8030	Forward	P63C domain protein
11622	14126	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
21327	21046	Reverse	Antitoxin igA-2
21772	21320	Reverse	Toxin HigB-2
22577	21717	Reverse	BRO family, N-terminal domain
23273	23605	Forward	HIRAN domain protein
23936	24370	Forward	Lysozyme RrrD
26446	25490	Reverse	Tyrosine recombinase XerC
26799	26485	Reverse	Helix-turn-helix domain protein
32879	32544	Reverse	LytTr DNA-binding domain protein
33502	32876	Reverse	YqaJ-like viral recombinase domain protein
34258	33506	Reverse	ERF superfamily protein
39863	39198	Reverse	putative HTH-type transcriptional regulator
40153	40656	Forward	Bacteriophage CII protein
40660	41256	Forward	T5orf172 domain protein

Table 30: Functional genes identified for vB_Pae_CF26a.


vB_Pae_CF26a
41525 bp
0.00kb
6.83kb
13.67kb
20.50kb
27.33kb
34.17kb
Bacteriophage CII protein
TSorf172 domain protein
hypothetical protein
Phage Terminase
hypothetical protein
Phage portal protein
cipP
Phage capsid family protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage headtail joining protein
P63C domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Lambda phage tail tapemeasure protein Ta
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Pae_CF26b genome size was 37.6 kb, the phage was assembled using SPAdes and has a mean coverage of 105. The genome map of this phage is shown in Figure 31. The total number of genes identified was 58 of which 10 were identified with putative functions these can be seen in Table 31. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF26b had a 64.4 % G-C content. No tRNA gene was identified however; Mu-like phage structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
1124	411	Reverse	HTH-type transcriptional regulator PrtR
1294	1650	Forward	DNA-binding transcriptional regulator Nlp
2352	4346	Forward	Mu DNA-binding domain protein
6559	7077	Forward	Bacteriophage Mu Gam like protein
9881	10330	Forward	Mor transcription activator family protein
16799	18085	Forward	Phage Mu protein F like protein
18085	18552	Forward	Phage virion morphogenesis family protein
19571	20677	Forward	Mu-like prophage I protein
21113	22021	Forward	Mu-like prophage major head subunit gpT
25053	28613	Forward	Prophage tail length tape measure protein

Table 31: Functional genes identified for vB_Pae_CF26b.

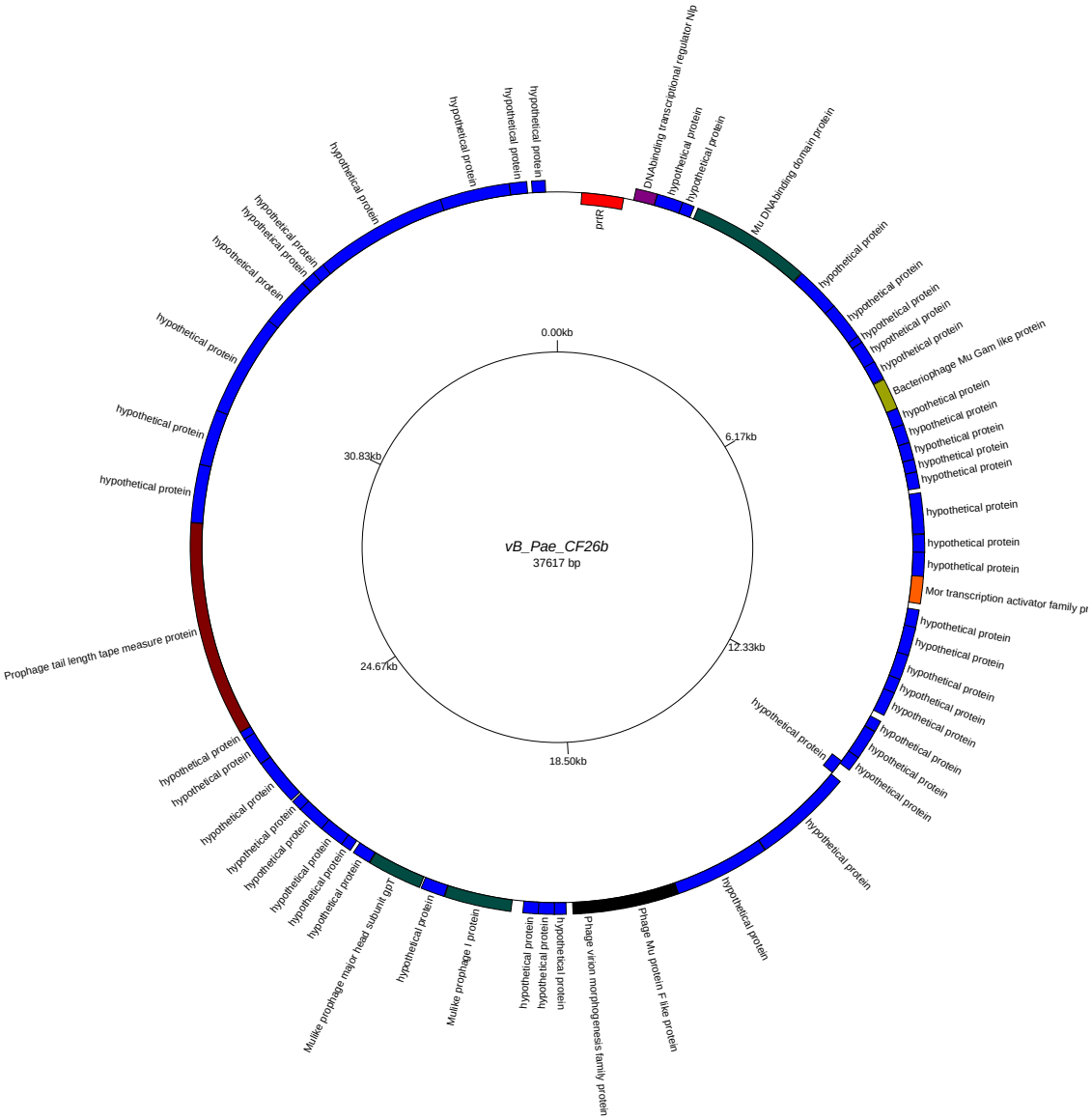


Figure 31: Genome map of vB_Pae_CF26b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF27a genome size was 39.7 kb, the phage was assembled using SPAdes and has a mean coverage of 146. The genome map of this phage is shown in Figure 32. The total number of genes identified was 56 of which 8 were identified with putative functions these can be seen in Table 32. Blastn on viruses only database showed a 7.1 kb / 7.6 kb match with *Pseudomonas* phage B3 (GenBank: NC_006548.1). The phage vB_Pae_CF27a had a 63.3 % G-C content. No tRNA gene was identified however; several genes associated with regulatory proteins were identified and an Integrase gene was identified.

Start	**End**	**Direction**	**Putative functional protein**
524	165	Reverse	Mor transcription activator family protein
8043	6259	Reverse	Integrase core domain protein
11052	11441	Forward	Helix-turn-helix domain protein
13313	13942	Forward	Soluble lytic murein transglycosylase precursor
16008	17711	Forward	Terminase-like family protein
19189	20439	Forward	Phage Mu protein F like protein
20436	21008	Forward	Phage virion morphogenesis family protein
27025	30792	Forward	Prophage tail length tape measure protein

Table 32: Functional genes identified for vB_Pae_CF27a.

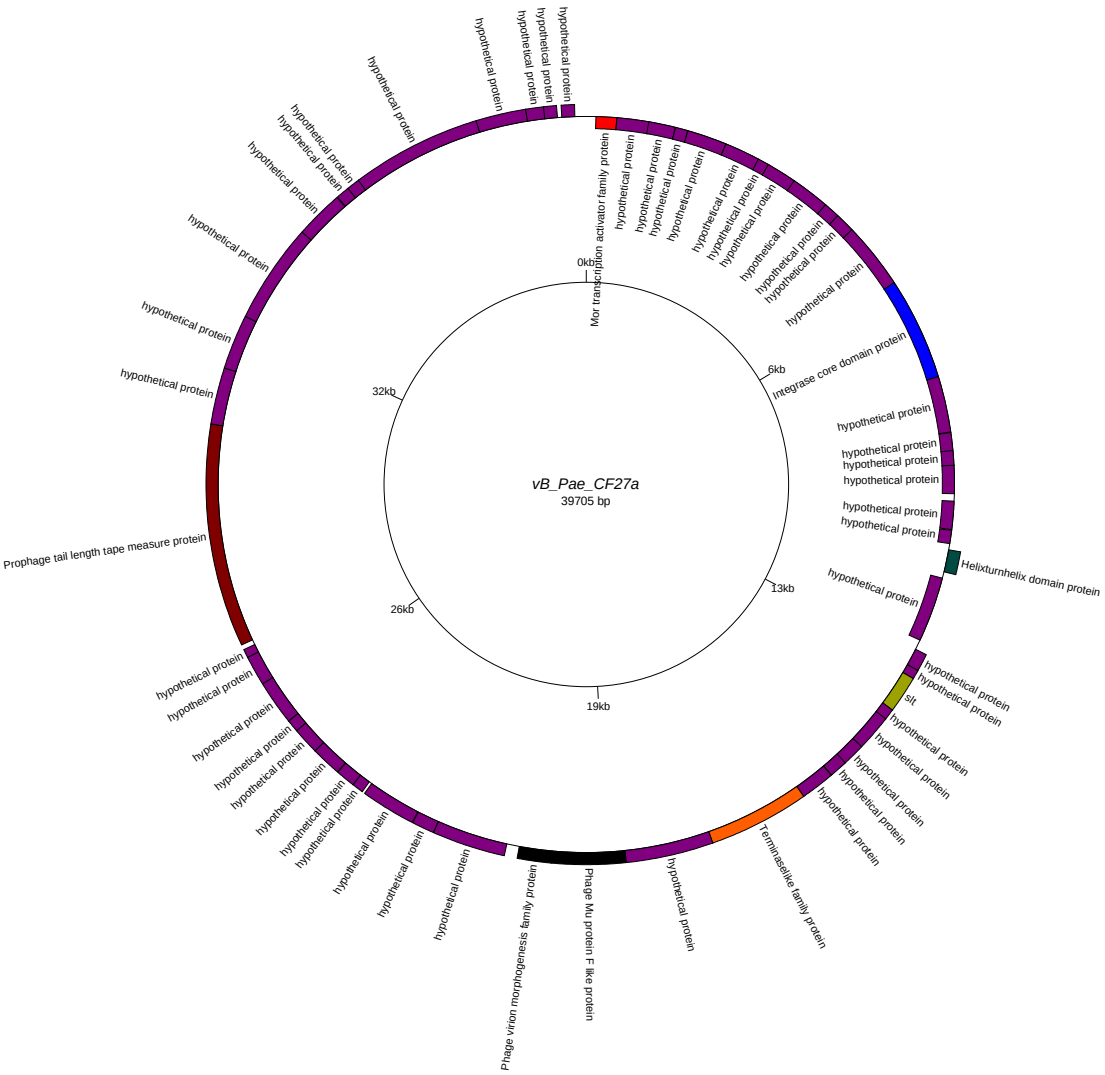


Figure 32: Genome map of vB_Pae_CF27a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF28a genome size was 41.1 kb, the phage was assembled using SPAdes and has a mean coverage of 1455. The genome map of this phage is shown in Figure 33. The total number of genes identified was 68 of which 14 were identified with putative functions these can be seen in Table 33. Blastn on viruses only database showed a 9.6 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF28a had a 61.6 % G-C content. No tRNA gene was identified however; genes associated with structural, lysogenic and regulatory proteins were identified, including an integrase putative protein.

Start	End	Direction	Putative functional protein
3307	2129	Reverse	Putative prophage phiRv2 integrase
5717	5487	Reverse	Arc-like DNA binding domain protein
6564	5797	Reverse	Phage regulatory protein Rha (Phage_pRha)
7969	7583	Reverse	Bacterial regulatory proteins, luxR family
9843	9052	Reverse	HTH-type transcriptional regulator PrtR
17566	17955	Forward	Phage antitermination protein Q
18494	19021	Forward	Phage regulatory protein Rha (Phage_pRha)
19106	19438	Forward	Phage holin family (Lysis protein S)
19435	20052	Forward	Chitinase class I
20288	20758	Forward	Bacteriophage lysis protein
21618	22163	Forward	Phage DNA packaging protein Nu1
22135	24099	Forward	Phage terminase large subunit (GpA)
24305	25951	Forward	Phage portal protein, lambda family
25923	28004	Forward	ATP-dependent Clp protease proteolytic subunit

Table 33: Functional genes identified for vB_Pae_CF28a.

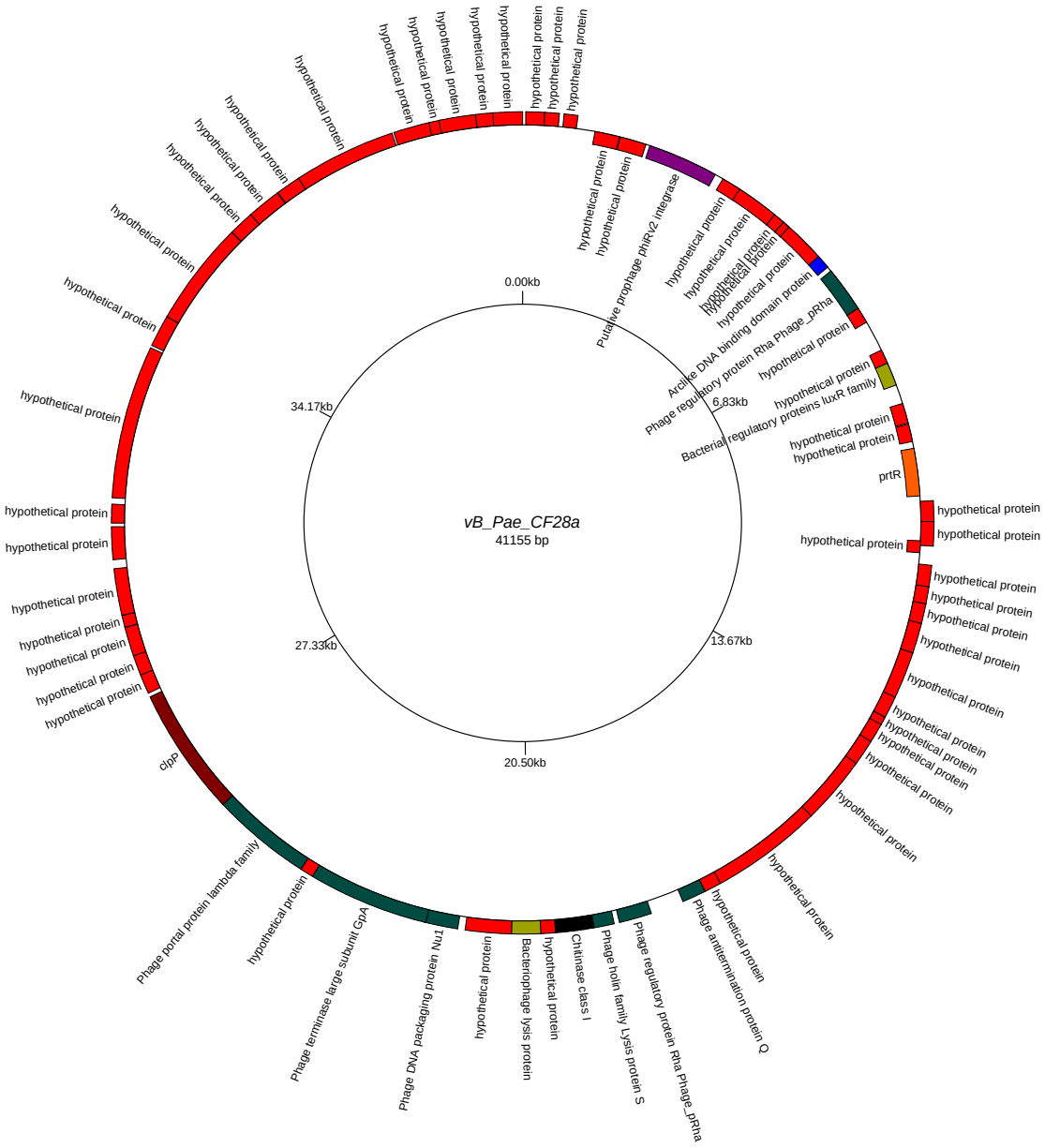


Figure 33: Genome map of vB_Pae_CF28a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The potential partial temperate phage vB_Pae_CF29a genome size was 52.8 kb, the phage was assembled using SPAdes with a mean coverage of 183. The genome map of this phage is shown in Figure 34. The total number of genes identified was 86 of which 18 were identified with putative functions these can be seen in Table 34. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_CF29a had a 59.1 % G-C content. A tRNA gene along with structural, lysogenic and regulatory proteins was identified. An Ig-like domain was also identified.

Start	End	Direction	Putative functional protein
923	327	Reverse	T5orf172 domain protein
1430	927	Reverse	Bacteriophage CII protein
1720	2385	Forward	putative HTH-type transcriptional regulator
6179	7243	Forward	IgA-specific serine endopeptidase autotransporter precursor
7251	7997	Forward	RecT family protein
7981	8601	Forward	YqaJ-like viral recombinase domain protein
8598	8933	Forward	LytTr DNA-binding domain protein
16695	15685	Reverse	site-specific tyrosine recombinase XerC
16856	16769	Reverse	tRNA-Ser(cga)
18360	17926	Reverse	Lysozyme RrrD
18963	19862	Forward	BRO family, N-terminal domain
22003	20357	Reverse	D-glucuronyl C5-epimerase C-terminus
29458	28898	Reverse	AP2 domain protein
34539	33544	Reverse	Bacterial Ig-like domain (group 2)
41154	39424	Reverse	Phage Mu protein F like protein
43805	42687	Reverse	Phage terminase large subunit
48630	47983	Reverse	Bacteriophage Lambda NinG protein
52429	51044	Reverse	Replicative DNA helicase

Table 34: Functional genes identified for vB_Pae_CF29a.

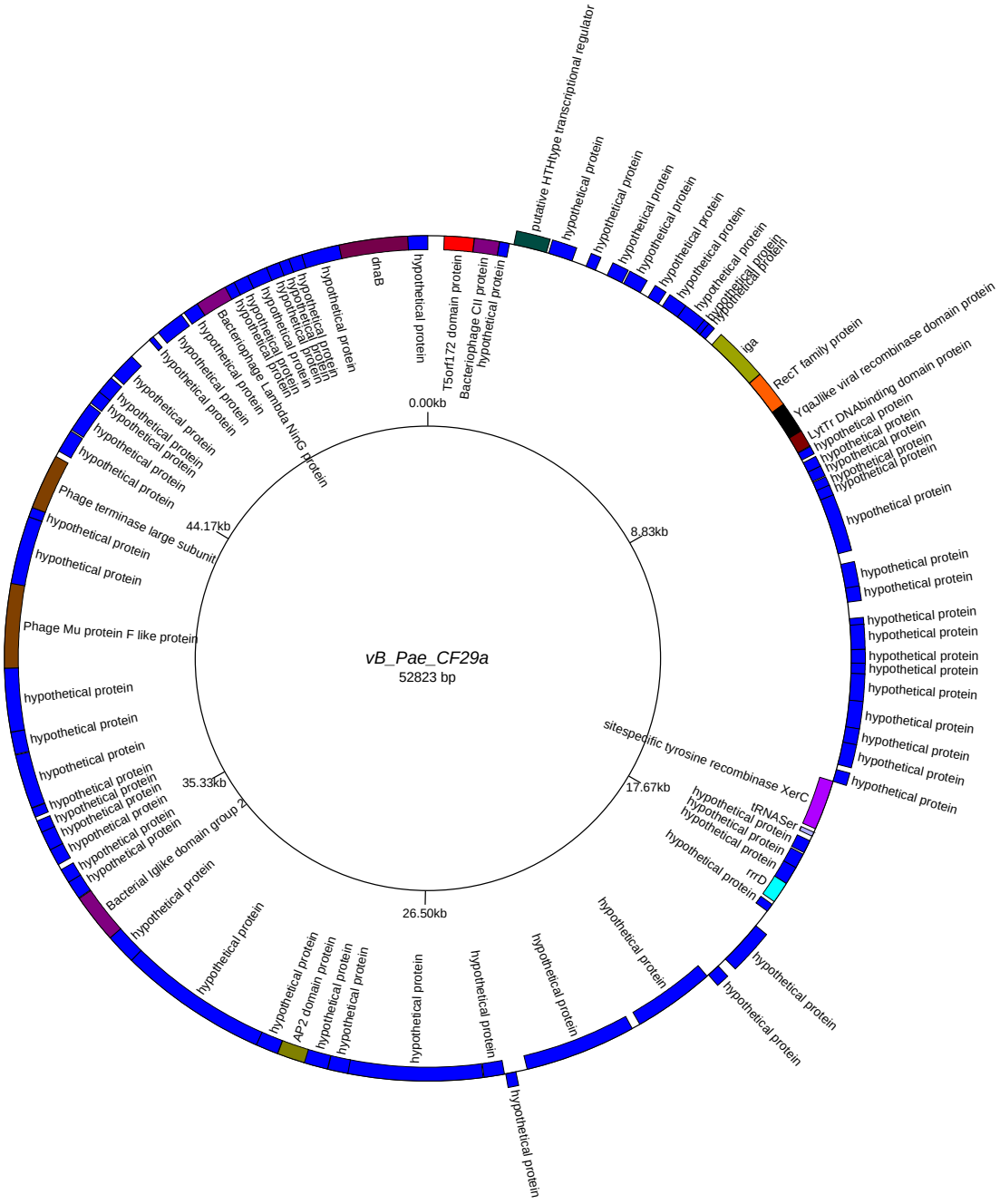
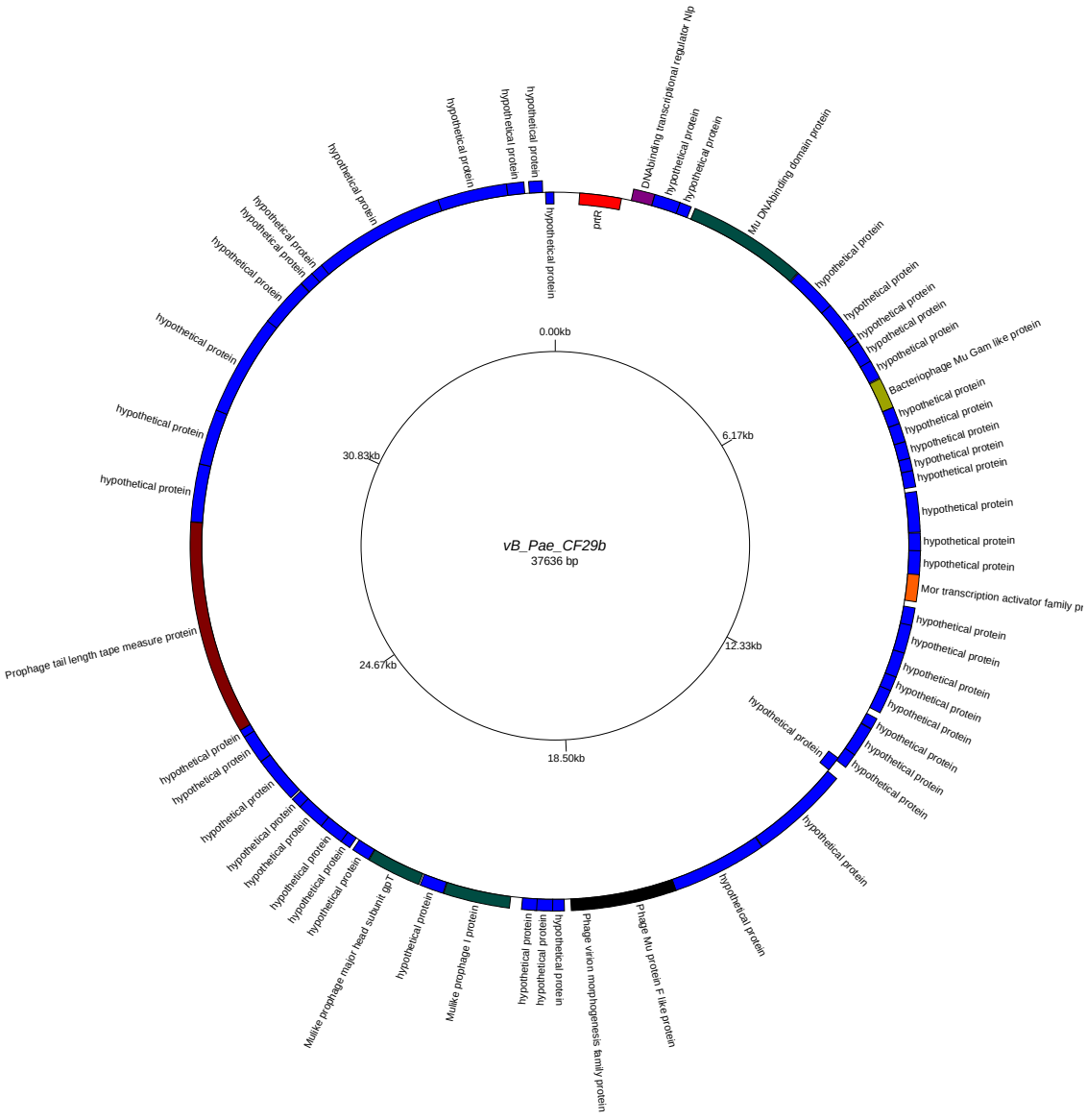


Figure 34: Genome map of vB_Pae_CF29a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF29b genome size was 37.6 kb, the phage was assembled using SPAdes and has a mean coverage of 162. The genome map of this phage is shown in Figure 35. The total number of genes identified was 59 of which 10 were identified with putative functions these can be seen in Table 35. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF29b had a 64.3 % G-C content. No tRNA gene was identified however; Mu-like phage structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
1127	414	Reverse	HTH-type transcriptional regulator PrtR
1297	1653	Forward	DNA-binding transcriptional regulator Nlp
2355	4349	Forward	Mu DNA-binding domain protein
6562	7080	Forward	Bacteriophage Mu Gam like protein
9884	10333	Forward	Mor transcription activator family protein
16802	18088	Forward	Phage Mu protein F like protein
18088	18555	Forward	Phage virion morphogenesis family protein
19574	20680	Forward	Mu-like prophage I protein
21116	22024	Forward	Mu-like prophage major head subunit gpT
25056	28616	Forward	Prophage tail length tape measure protein

Table 35: Functional genes identified for vB_Pae_CF29b.


vB_Pae_CF29b
37636 bp
0.00kb
6.17kb
12.33kb
18.50kb
24.67kb
30.83kb
DNAbinding transcriptional regulator Nlp
hypothetical protein
hypothetical protein
Mu DNAbinding domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Bacteriophage Mu Gam like protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Mor transcription activator family pr
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage Mu protein F like protein
Phage virion morphogenesis family protein
hypothetical protein
hypothetical protein
hypothetical protein
Mulike prophage I protein
hypothetical protein
Multie prophage major head subunit gp1
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Prophage tail length tape measure protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The potential partial temperate phage vB_Pae_CF30a genome size was 52.7 kb, the phage was assembled using SPAdes with a mean coverage of 28. The genome map of this phage is shown in Figure 36. The total number of genes identified was 86 of which 18 were identified with putative functions these can be seen in Table 36. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_CF30a had a 58.9 % G-C content. A tRNA gene along with structural, lysogenic and regulatory proteins was identified. An Ig-like domain was also identified.

Start	End	Direction	Putative functional protein
874	278	Reverse	T5orf172 domain protein
1381	878	Reverse	Bacteriophage CII protein
1671	2336	Forward	putative HTH-type transcriptional regulator
6130	7194	Forward	IgA-specific serine endopeptidase autotransporter precursor
7202	7948	Forward	RecT family protein
7932	8552	Forward	YqaJ-like viral recombinase domain protein
8549	8884	Forward	LytTr DNA-binding domain protein
16646	15636	Reverse	site-specific tyrosine recombinase XerC
16807	16720	Reverse	tRNA-Ser(cga)
18311	17877	Reverse	Lysozyme RrrD
18914	19813	Forward	BRO family, N-terminal domain
21954	20308	Reverse	D-glucuronyl C5-epimerase C-terminus
29409	28849	Reverse	AP2 domain protein
34490	33495	Reverse	Bacterial Ig-like domain (group 2)
41105	39375	Reverse	Phage Mu protein F like protein
43756	42638	Reverse	Phage terminase large subunit
48581	47934	Reverse	Bacteriophage Lambda NinG protein
52380	50995	Reverse	Replicative DNA helicase

Table 36: Functional genes identified for vB_Pae_CF30a.

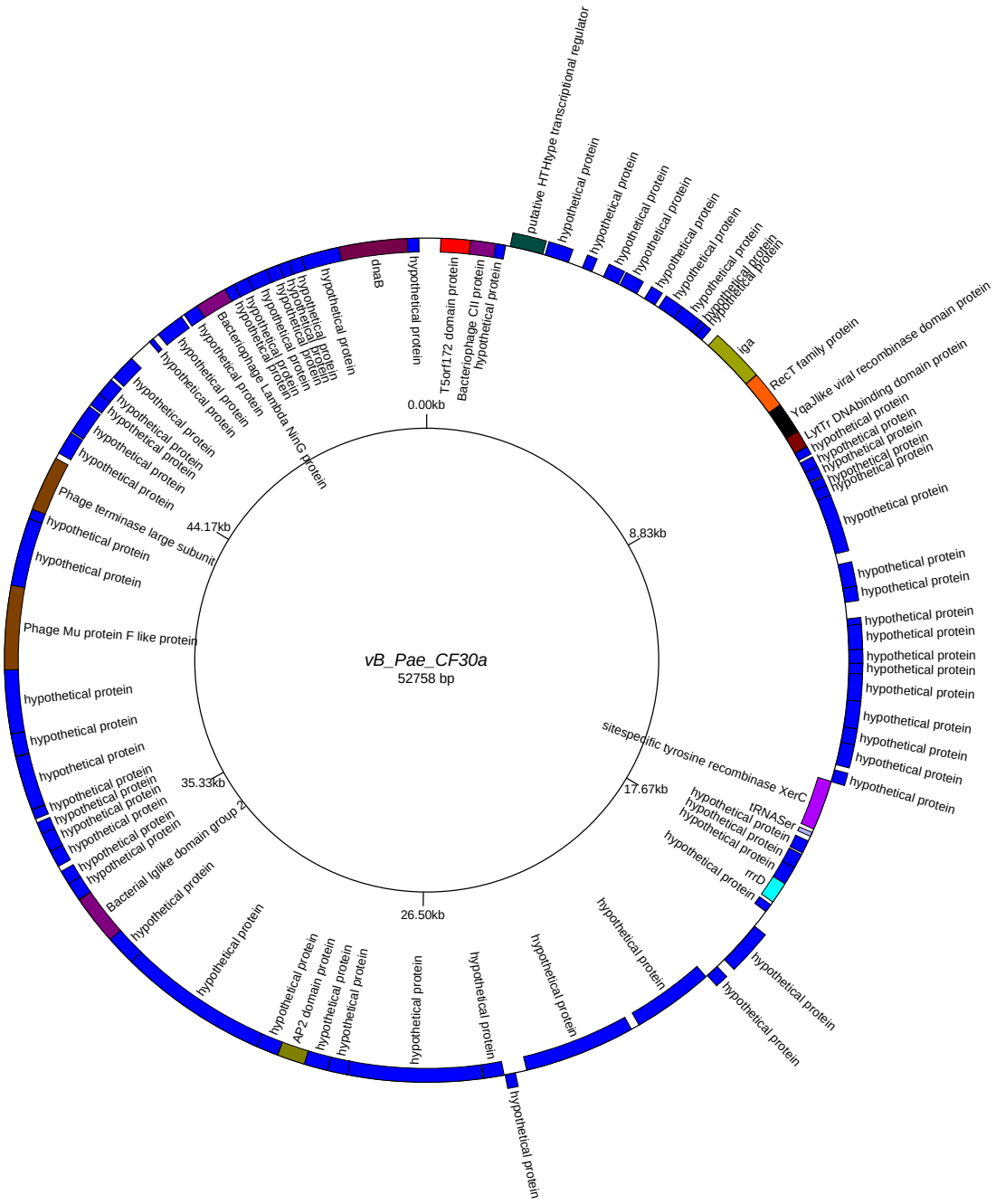


Figure 36: Genome map of vB_Pae_CF30a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF30b genome size was 37.7 kb, the phage was assembled using SPAdes and extended using PriceTI. The genome map of this phage is shown in Figure 37. The total number of genes identified was 59 of which 11 were identified with putative functions these can be seen in Table 37. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF30b had a 64.4 % G-C content. No tRNA gene was identified however; Mu-like phage structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
1153	440	Reverse	HTH-type transcriptional regulator PrtR
1323	1679	Forward	DNA-binding transcriptional regulator Nlp
2381	4375	Forward	Mu DNA-binding domain protein
6588	7106	Forward	Bacteriophage Mu Gam like protein
9910	10359	Forward	Mor transcription activator family protein
16828	18114	Forward	Phage Mu protein F like protein
18114	18581	Forward	Phage virion morphogenesis family protein
19600	20706	Forward	Mu-like prophage I protein
21142	22050	Forward	Mu-like prophage major head subunit gpT
25082	28642	Forward	Prophage tail length tape measure protein
37731	37561	Reverse	LysR substrate binding domain protein

Table 37: Functional genes identified for vB_Pae_CF30b.

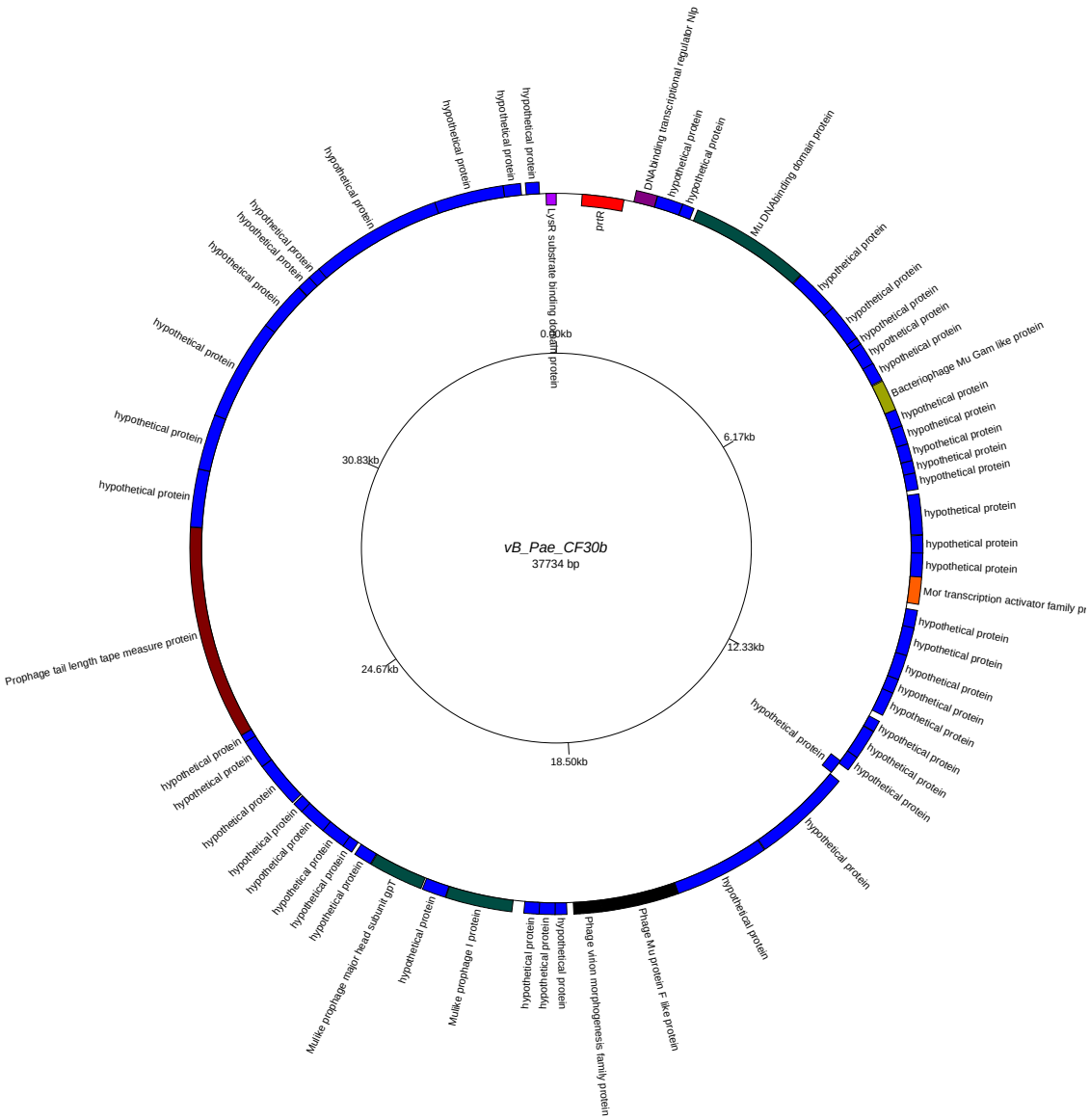


Figure 37: Genome map of vB_Pae_CF30b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF30c genome size was 32.1 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 38. The total number of genes identified was 48 of which 8 were identified with putative functions these can be seen in Table 38. Blastn on viruses only database showed a 2.5 kb / 2.6 kb match with *Pseudomonas* phage B3 (GenBank: NC_006548.1). The phage vB_Pae_CF30c had a 63.4 % G-C content. No tRNA gene was identified however; several genes associated with regulatory proteins were identified and an Integrase gene was identified.

Start	End	Direction	Putative functional protein
466	107	Reverse	Mor transcription activator family protein
7985	6201	Reverse	Integrase core domain protein
10994	11383	Reverse	Helix-turn-helix domain protein
13254	13883	Forward	Soluble lytic murein transglycosylase precursor
15949	17652	Forward	Terminase-like family protein
19130	20380	Forward	Phage Mu protein F like protein
20377	20949	Forward	Phage virion morphogenesis family protein
26966	30733	Forward	Prophage tail length tape measure protein

Table 38: Functional genes identified for vB_Pae_CF30c.

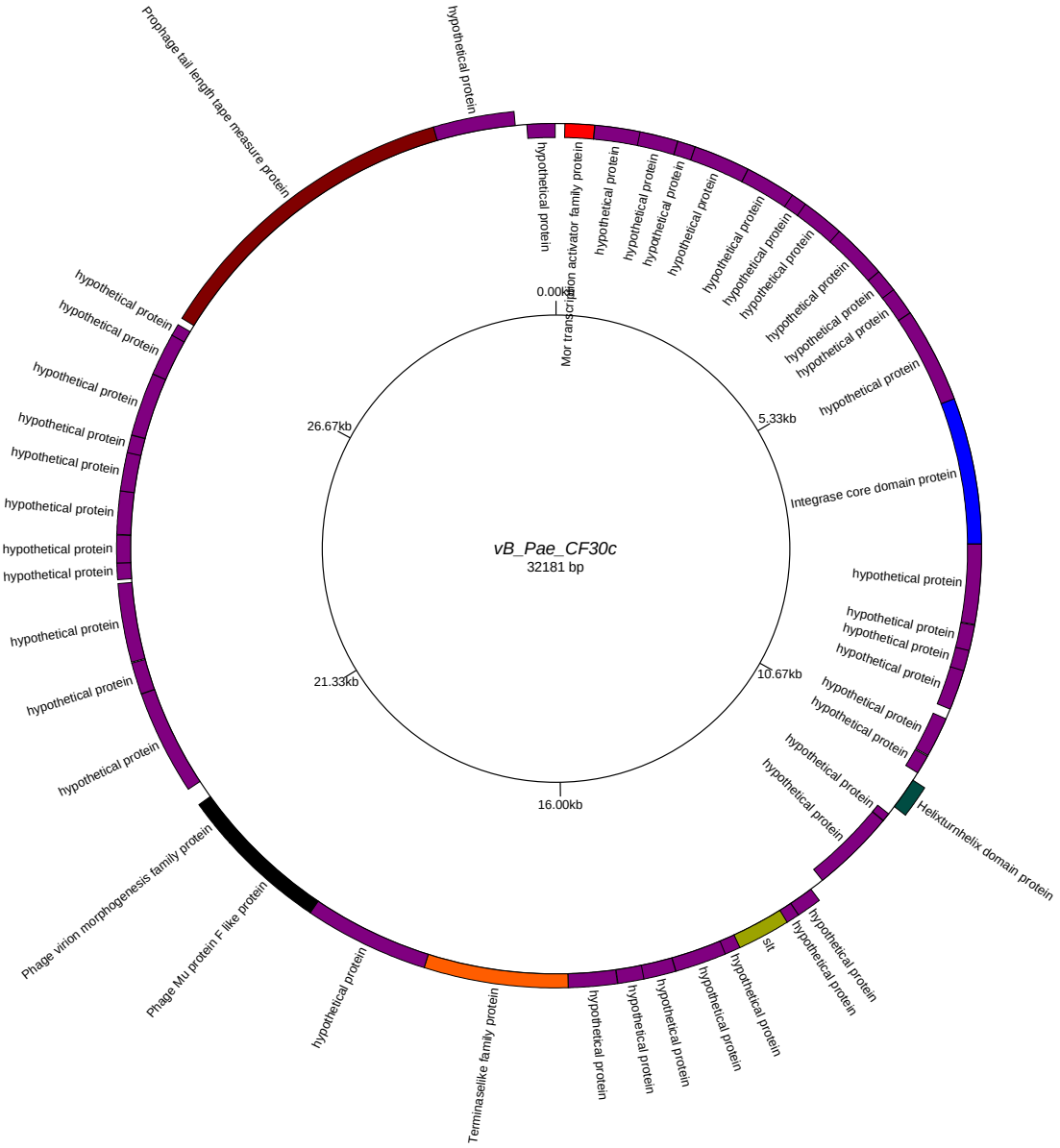


Figure 38: Genome map of vB_Pae_CF30c produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF31a genome size was 37.6 kb, the phage was assembled using SPAdes and has a mean coverage of 329. The genome map of this phage is shown in Figure 39. The total number of genes identified was 58 of which 10 were identified with putative functions these can be seen in Table 39. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF31a had a 64.4 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
1126	413	Reverse	HTH-type transcriptional regulator PrtR
1296	1652	Forward	DNA-binding transcriptional regulator Nlp
2354	4348	Forward	Mu DNA-binding domain protein
6561	7079	Forward	Bacteriophage Mu Gam like protein
9883	10332	Forward	Mor transcription activator family protein
16801	18087	Forward	Phage Mu protein F like protein
18087	18554	Forward	Phage virion morphogenesis family protein
19573	20679	Forward	Mu-like prophage I protein
21115	22023	Forward	Mu-like prophage major head subunit gpT
25055	28615	Forward	Prophage tail length tape measure protein

Table 39: Functional genes identified for vB_Pae_CF31a.

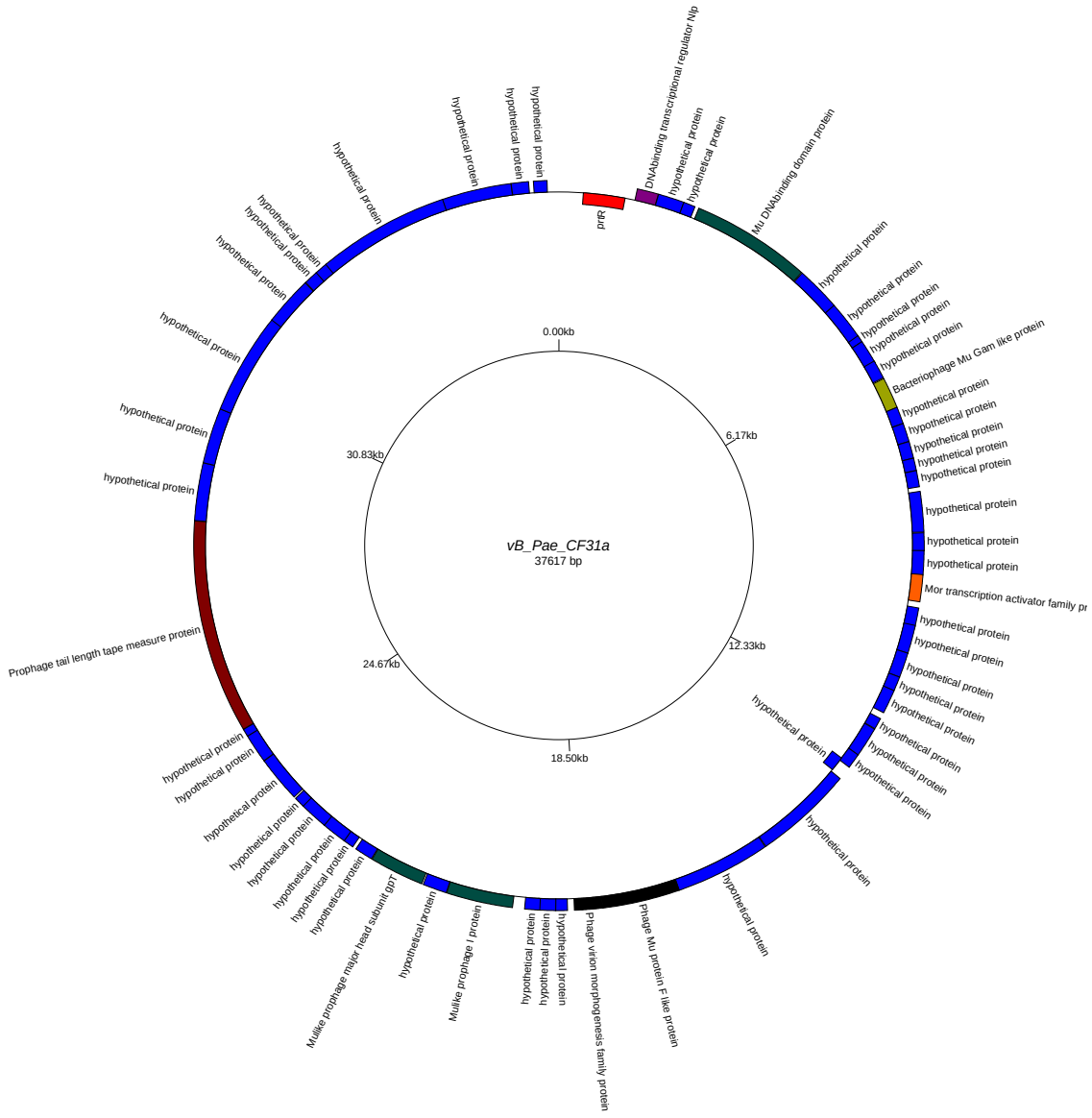


Figure 39: Genome map of vB_Pae_CF31a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF31b genome size was 42.4 kb, the phage was assembled using SPAdes with a mean coverage of 195. The genome map of this phage is shown in Figure 40. The total number of genes identified was 66 of which 8 were identified with putative functions these can be seen in Table 40. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_CF31b had a 58.9 % G-C content. A tRNA gene was identified along with structural, regulatory and lysogenic proteins. An Ig-like domain was also identified.

Start	End	Direction	Putative functional protein
843	247	Reverse	T5orf172 domain protein
1350	847	Reverse	Bacteriophage CII protein
1640	2305	Forward	putative HTH-type transcriptional regulator
6099	7163	Forward	IgA-specific serine endopeptidase autotransporter precursor
7171	7917	Forward	RecT family protein
7901	8521	Forward	YqaJ-like viral recombinase domain protein
8518	8853	Forward	LytTr DNA-binding domain protein
16615	15605	Reverse	site-specific tyrosine recombinase XerC
16776	16689	Reverse	tRNA-Ser(cga)
18280	17846	Reverse	Lysozyme RrrD
18883	19782	Forward	BRO family, N-terminal domain
21923	20277	Reverse	D-glucuronyl C5-epimerase C-terminus
29378	28818	Reverse	AP2 domain protein
34459	33464	Reverse	Bacterial Ig-like domain (group 2)
41074	39344	Reverse	Phage Mu protein F like protein

Table 40: Functional genes identified for vB_Pae_CF31b.

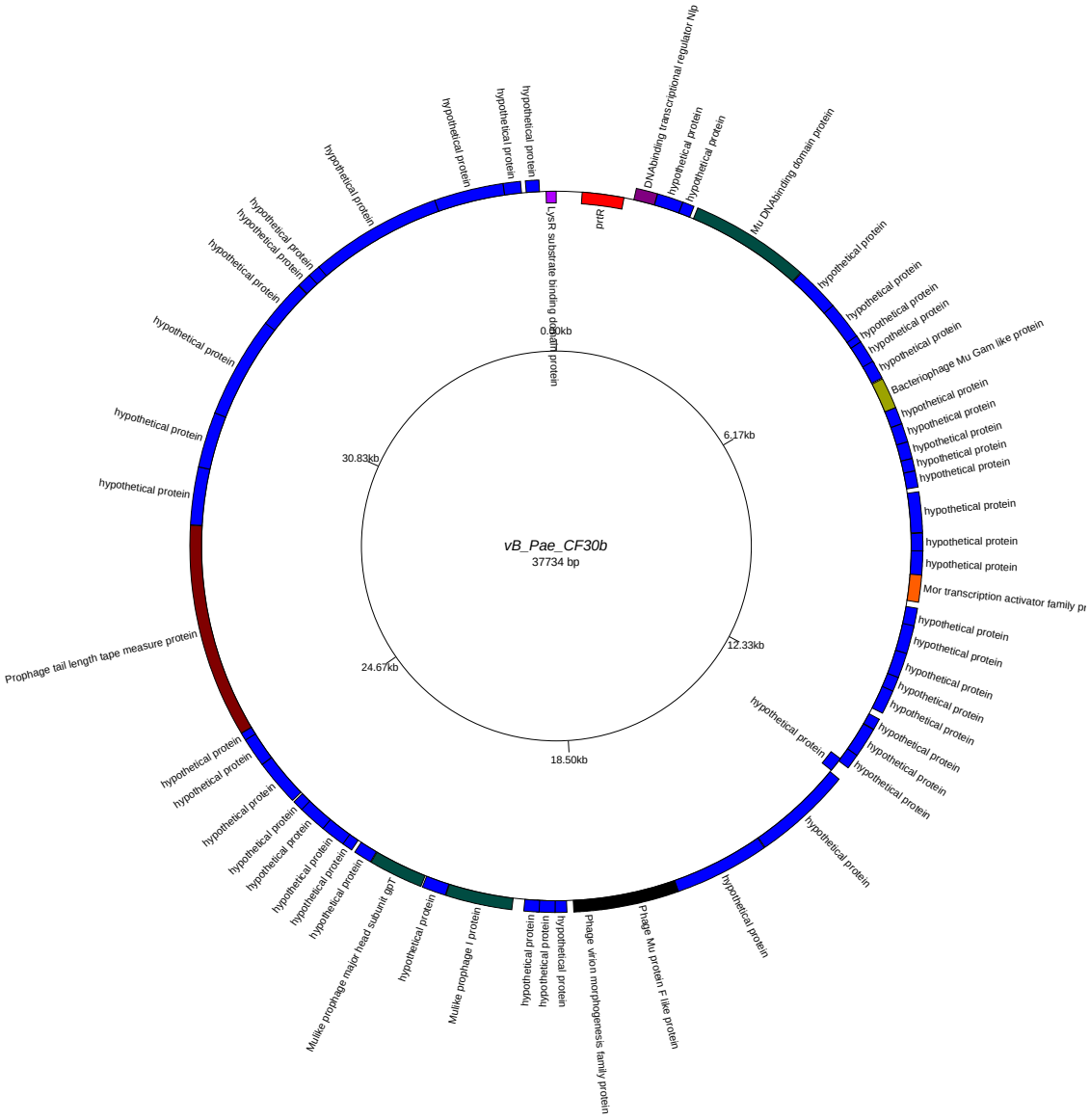


Figure 40: Genome map of vB_Pae_CF31b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF31c genome size was 38.3 kb, the phage was assembled using SPAdes and has a mean coverage of 35. The genome map of this phage is shown in Figure 41. The total number of genes identified was 58 of which 26 were identified with putative functions these can be seen in Table 41. Blastn on viruses only database showed a 1.9 kb / 1.9 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF31c had a 61.3 % G-C content. No tRNA gene was identified however; structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
503	2182	Forward	Phage Terminase
2185	3408	Forward	Phage portal protein
3392	4036	Forward	Caudovirus prohead protease
4033	5247	Forward	Phage capsid family protein
5502	5825	Forward	Phage gp6-like head-tail connector protein
5825	6151	Forward	Phage head-tail joining protein
8523	11798	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
11798	12136	Forward	Phage minor tail protein
12133	12879	Forward	Phage minor tail protein L
12882	13640	Forward	NlpC/P60 family protein
14391	14963	Forward	Bacteriophage lambda tail assembly protein I
15020	18685	Forward	Carbohydrate binding domain protein
20273	19290	Reverse	Transposase DDE domain protein
24981	24751	Reverse	Arc-like DNA binding domain protein
25841	25056	Reverse	BRO family, N-terminal domain
26539	26153	Reverse	Bacterial regulatory proteins, luxR family
27347	26688	Reverse	LexA repressor
27457	27762	Forward	Helix-turn-helix
31718	32533	Forward	DNA replication protein DnaC
32530	33927	Forward	Replicative DNA helicase
34199	34588	Forward	Phage antitermination protein Q
35363	35893	Forward	Phage regulatory protein Rha (Phage_pRha)

35978	36310	Forward	Phage holin family (Lysis protein S)
36307	36924	Forward	Chitinase class I
36924	37436	Forward	Bacteriophage lysis protein
37631	37960	Forward	HNH endonuclease

Table 41: Functional genes identified for vB_Pae_CF31c.

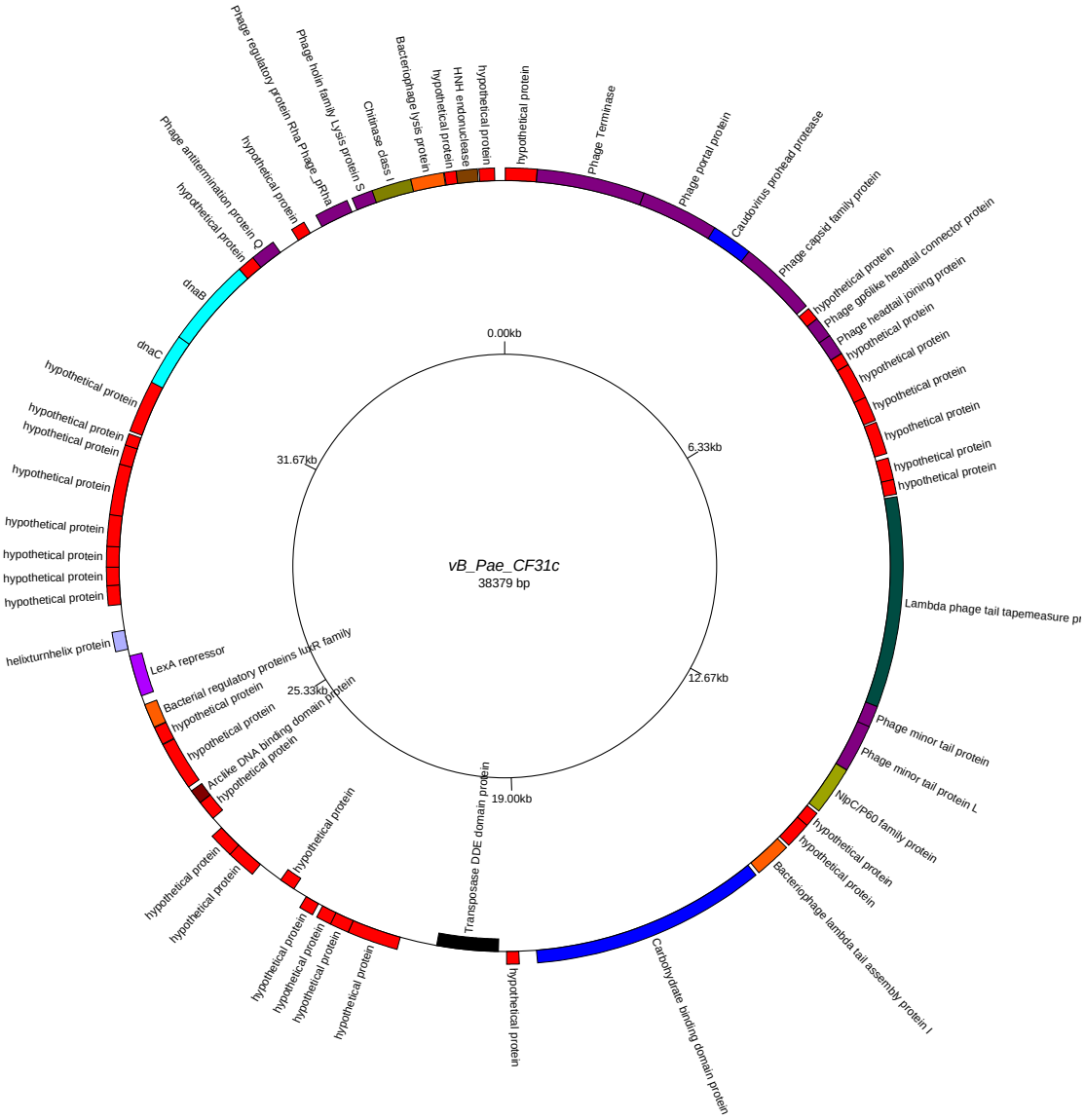


Figure 41: Genome map of vB_Pae_CF31c produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF32a genome size was 50.2 kb, the phage was assembled using SPAdes and has a mean coverage of 180. The genome map of this phage is shown in Figure 42. The total number of genes identified was 80 of which 14 were identified with putative functions these can be seen in Table 42. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF32a had a 63.2 % G-C content. No tRNA gene was identified however; genes associated with structural, lysogenic and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
12564	9004	Reverse	Prophage tail length tape measure protein
16504	15596	Reverse	Mu-like prophage major head subunit gpT
18046	16940	Reverse	Mu-like prophage I protein
19532	19065	Reverse	Phage virion morphogenesis family protein
20818	19532	Reverse	Phage Mu protein F like protein
27736	27287	Reverse	Mor transcription activator family protein
31058	30540	Reverse	Bacteriophage Mu Gam like protein
35265	33271	Reverse	Mu DNA-binding domain protein
36323	35967	Reverse	DNA-binding transcriptional regulator Nlp
36493	37206	Forward	HTH-type transcriptional regulator PrtR
38741	37542	Reverse	Phage Mu protein F like protein
41392	40274	Reverse	Phage terminase large subunit
46217	45570	Reverse	Bacteriophage Lambda NinG protein
50016	48631	Reverse	Replicative DNA helicase

Table 41: Functional genes identified for vB_Pae_CF32a.

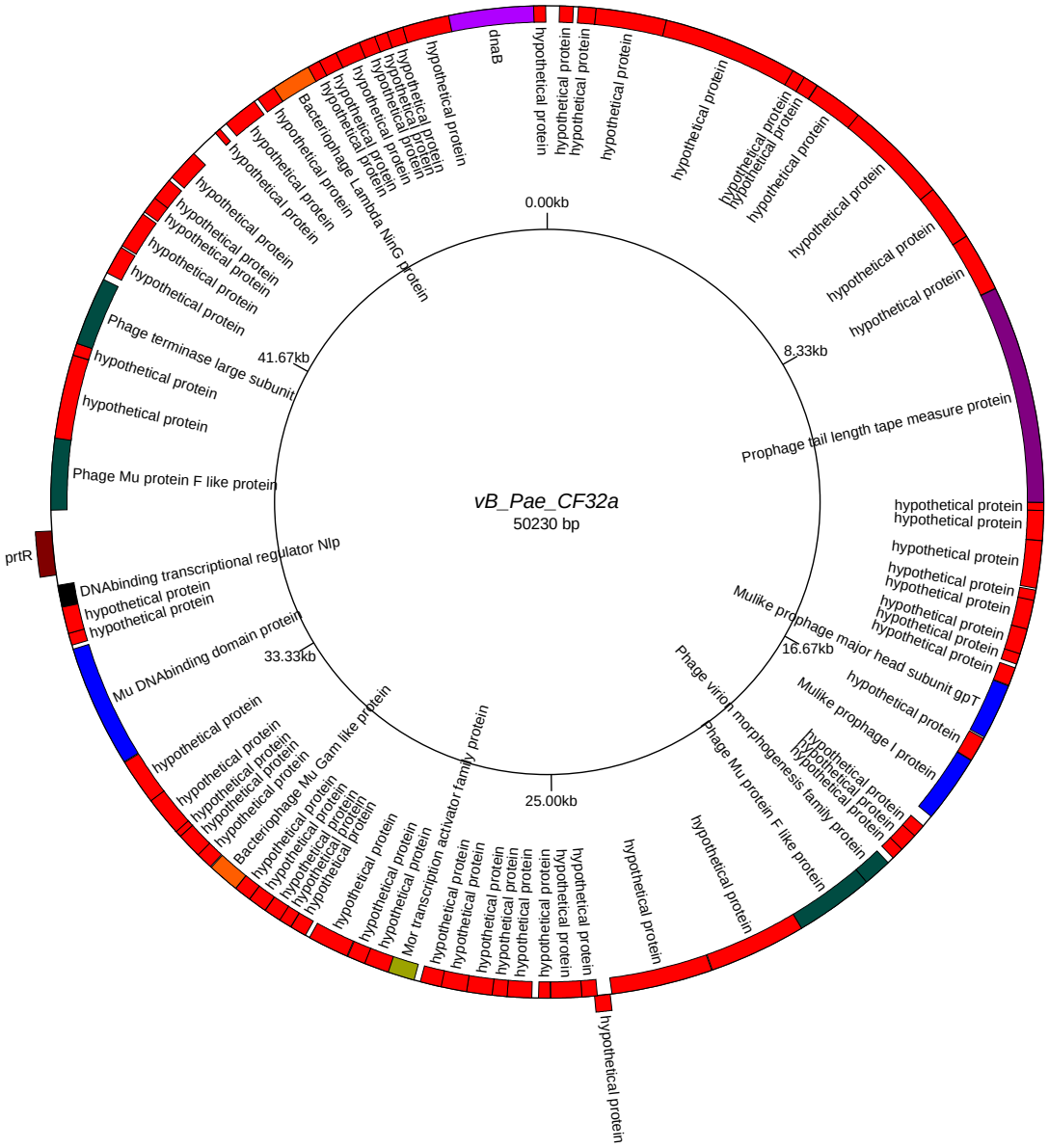


Figure 42: Genome map of vB_Pae_CF32a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The potential partial temperate phage vB_Pae_CF32b genome size was 29.1 kb, the phage was assembled using SPAdes with a mean coverage of 180. The genome map of this phage is shown in Figure 43. The total number of genes identified was 42 of which 7 were identified with putative functions these can be seen in Table 43. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_CF32b had a 58.8 % G-C content. A tRNA gene along with recombinase and regulatory proteins was identified. An Ig-like domain was also identified.

Start	End	Direction	Putative functional protein
5686	4676	Reverse	site-specific tyrosine recombinase XerC
5847	5760	Reverse	tRNA-Ser(cga)
7351	6917	Reverse	Lysozyme RrrD
7954	8853	Forward	BRO family, N-terminal domain
10994	9348	Reverse	D-glucuronyl C5-epimerase C-terminus
18449	17889	Reverse	AP2 domain protein
23530	22535	Reverse	Bacterial Ig-like domain (group 2)

Table 43: Functional genes identified for vB_Pae_CF32b.

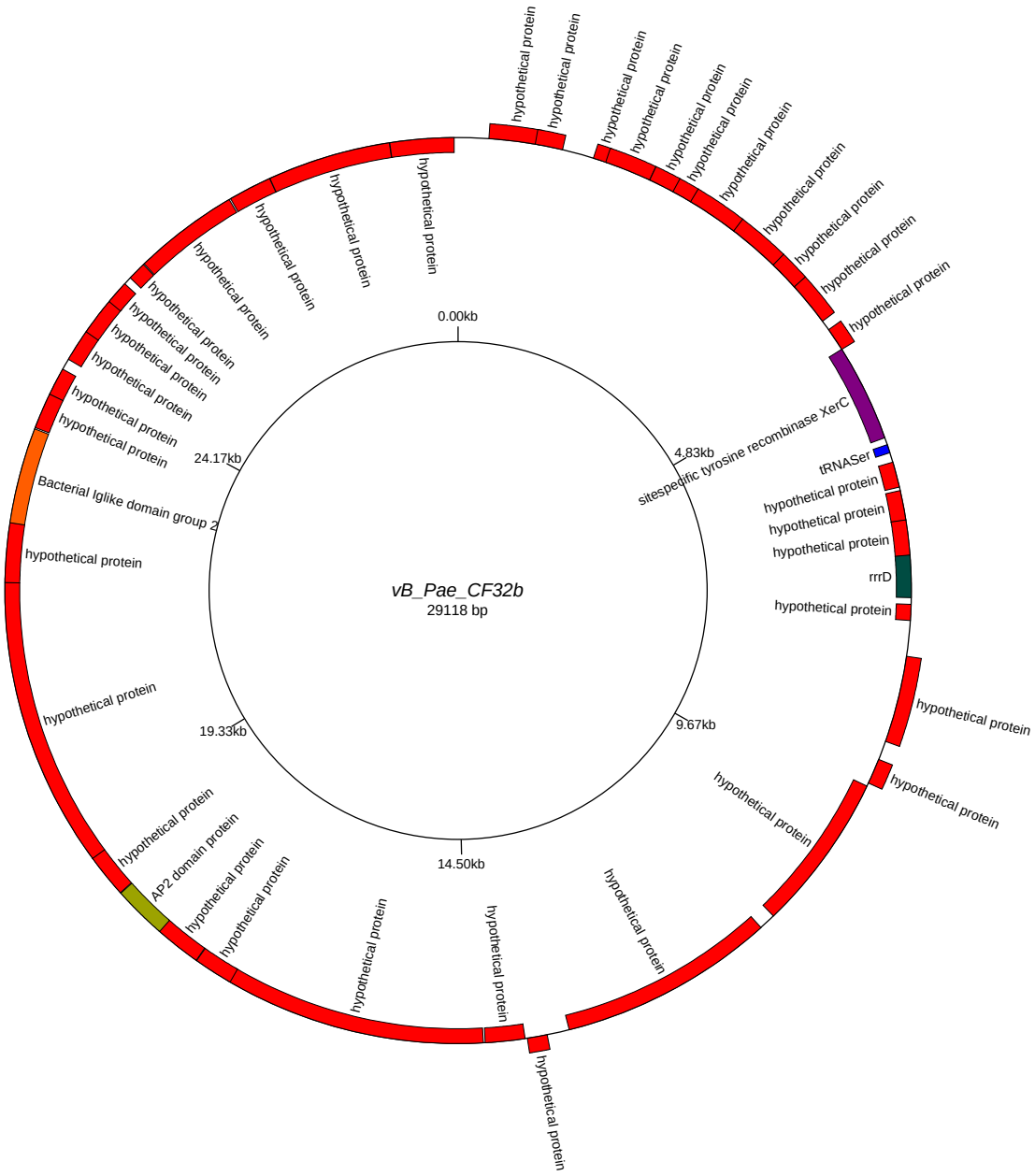


Figure 43: Genome map of vB_Pae_CF32b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF33a genome size was 39.6 kb, the phage was assembled using SPAdes and has a mean coverage of 720. The genome map of this phage is shown in Figure 44. The total number of genes identified was 56 of which 8 were identified with putative functions these can be seen in Table 44. Blastn on viruses only database showed a 7.1 kb / 7.6 kb match with *Pseudomonas* phage B3 (GenBank: NC_006548.1). The phage vB_Pae_CF33a had a 63.5 % G-C content. No tRNA gene was identified however; several genes associated with regulatory proteins and an Integrase protein was identified.

Start	End	Direction	Putative functional protein
523	164	Reverse	Mor transcription activator family protein
8042	6258	Reverse	Integrase core domain protein
11051	11440	Forward	Helix-turn-helix domain protein
13312	13941	Forward	Soluble lytic murein transglycosylase precursor
16007	17710	Forward	Terminase-like family protein
19188	20438	Forward	Phage Mu protein F like protein
20435	21007	Forward	Phage virion morphogenesis family protein
27024	30791	Forward	Prophage tail length tape measure protein

Table 44: Functional genes identified for vB_Pae_CF33a.

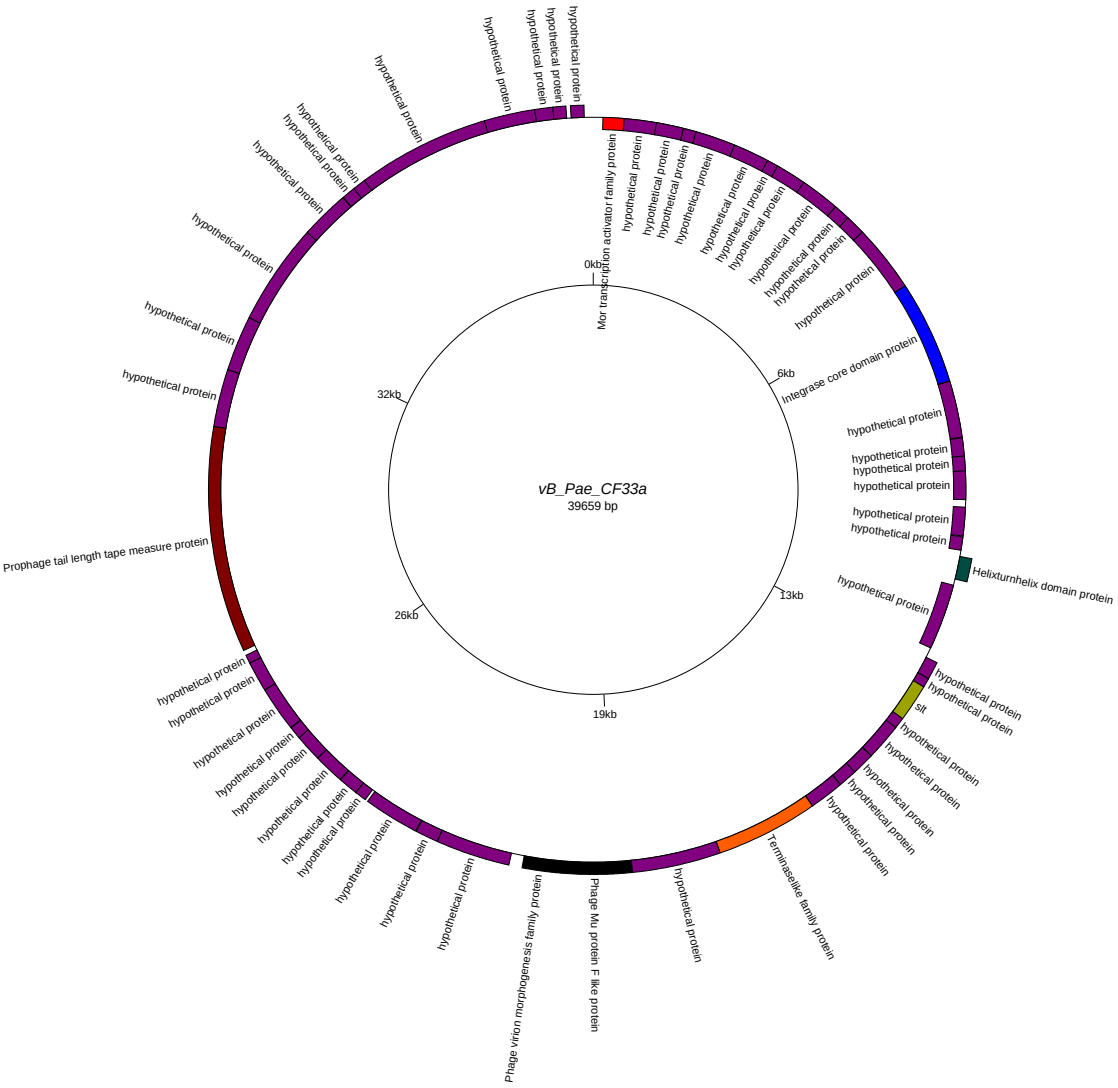
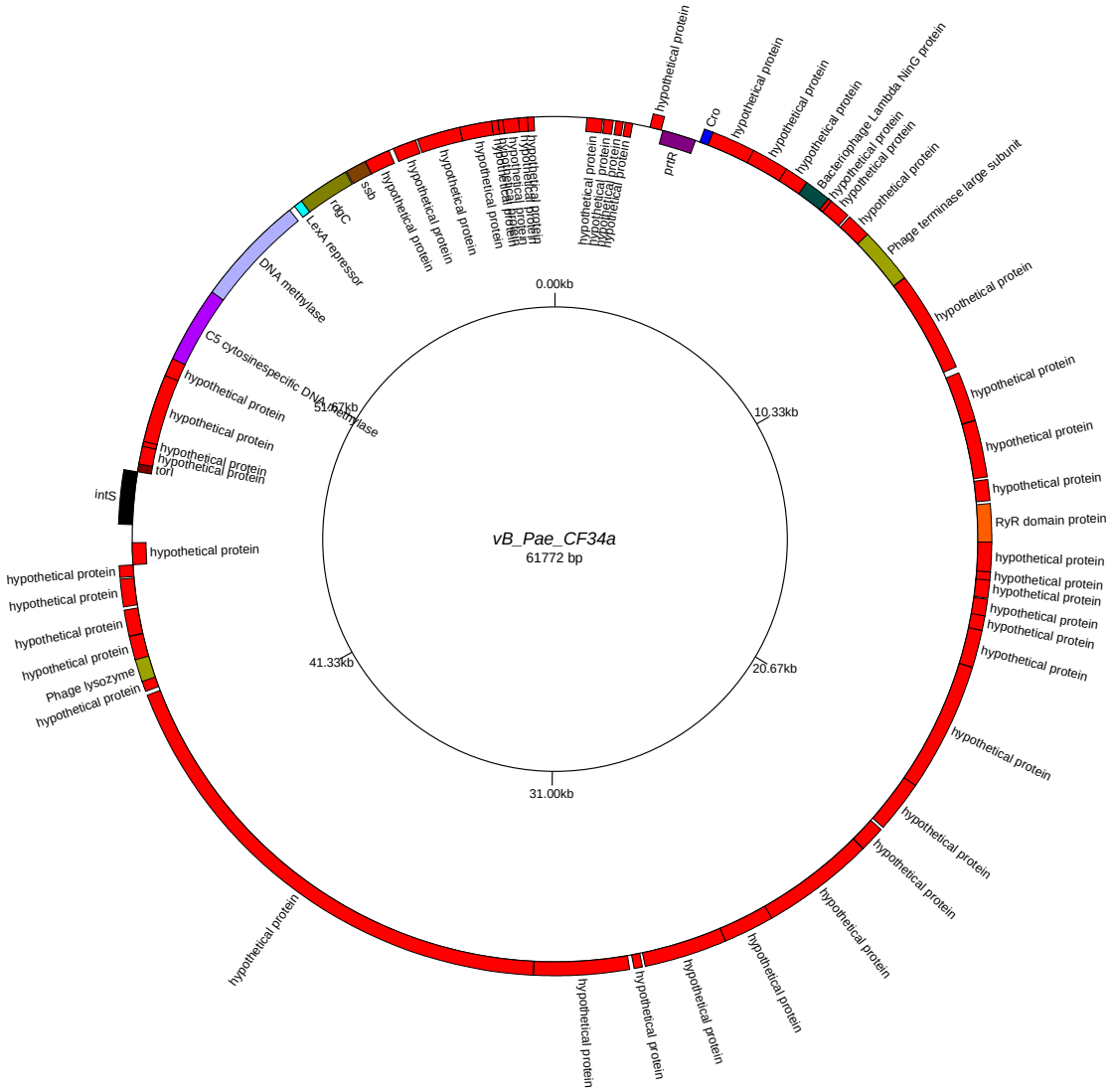


Figure 44: Genome map of vB_Pae_CF33a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF34a genome size was 61.7 kb, the phage was assembled using SPAdes and has a mean coverage of 15. The genome map of this phage is shown in Figure 45. The total number of genes identified was 62 of which 14 were identified with putative functions these can be seen in Table 45. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66 (GenBank: KC262634.1). The phage vB_Pae_CF34a had a 63.7 % G-C content. No tRNA gene was identified however; regulatory and lysogenic proteins were identified. A CPS-53 integrase like gene and *Cro* gene were identified.

Start	End	Direction	Putative functional protein
1112	741	Reverse	Carbon storage regulator homolog
3328	2531	Reverse	HTH-type transcriptional regulator PrtR
3436	3636	Forward	Cro
6042	6656	Forward	Bacteriophage Lambda NinG protein
7840	9123	Forward	Phage terminase large subunit
14639	15511	Forward	RyR domain protein
43063	43602	Forward	Phage lysozyme
46668	47903	Forward	Putative prophage CPS-53 integrase
48137	47904	Reverse	Response regulator inhibitor for tor operon
52474	50666	Reverse	C-5 cytosine-specific DNA methylase
55101	52471	Reverse	DNA methylase
55455	55246	Reverse	LexA repressor
56672	55452	Reverse	Recombination-associated protein RdgC
57196	56699	Reverse	Single-stranded DNA-binding protein

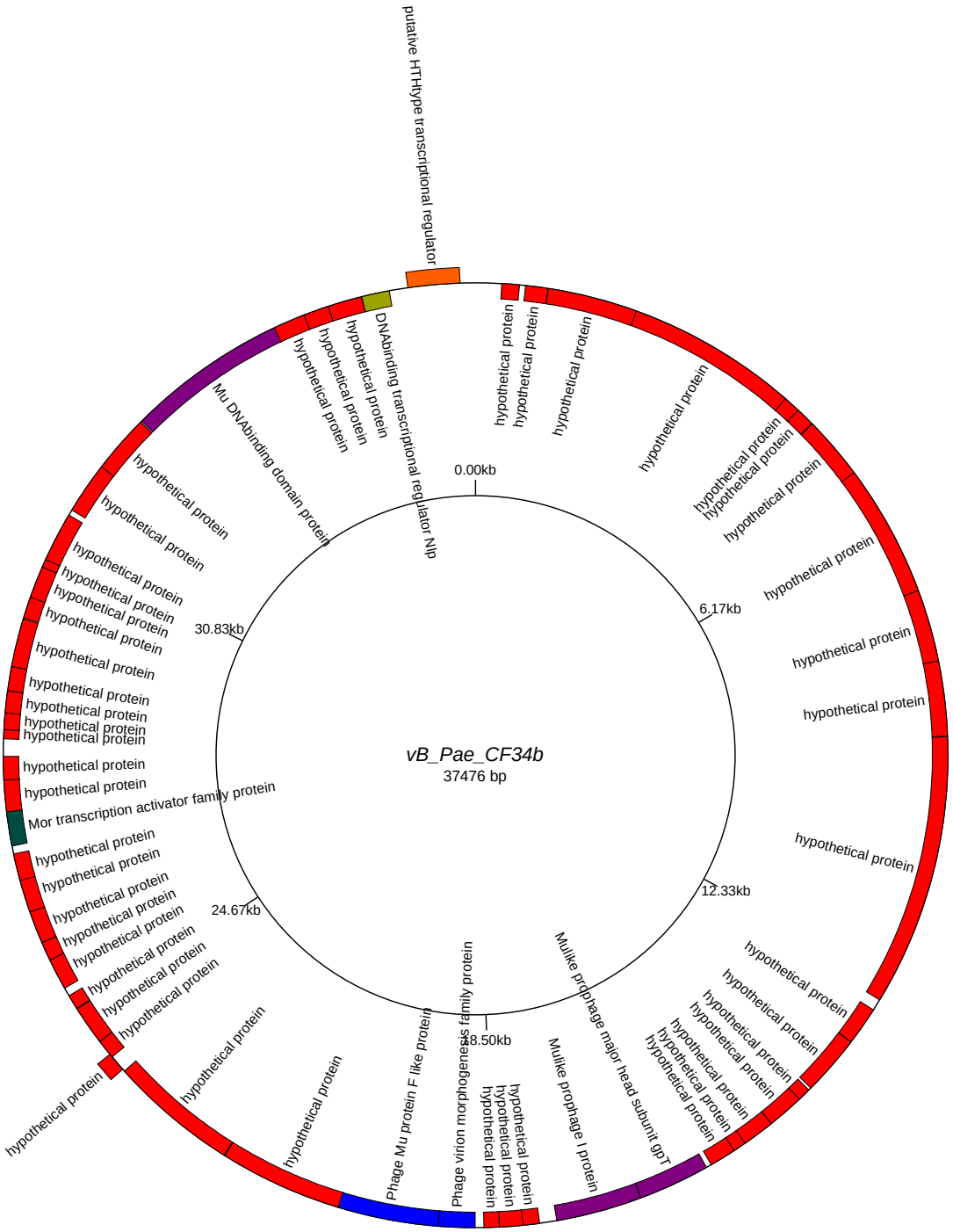
Table 45: Functional genes identified for vB_Pae_CF34a.


vB_Pae_CF34a
61772 bp
0.00kb
10.33kb
20.67kb
31.00kb
41.33kb
51.67kb
hypothetical protein
pntR
Cro
hypothetical protein
hypothetical protein
hypothetical protein
Bacteriophage Lambda NinG protein
hypothetical protein
hypothetical protein
Phage terminase large subunit
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
RyR domain protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Pae_CF34b genome size was 37.4 kb, the phage was assembled using SPAdes and has a mean coverage of 259. The genome map of this phage is shown in Figure 46. The total number of genes identified was 56 of which 8 were identified with putative functions these can be seen in Table 46. Blastn on viruses only database showed a 27.1 kb / 27.8 kb match with *Pseudomonas* phage JBD24 (GenBank: NC_020203.1). The phage vB_Pae_CF34b had a 64.1 % G-C content. No tRNA gene was identified however; several Mu like proteins were identified associated with morphogenesis and structural proteins.

Start	End	Direction	Putative functional protein
16605	15691	Reverse	Mu-like prophage major head subunit gpT
17706	16609	Reverse	Mu-like prophage I protein
19211	18744	Reverse	Phage virion morphogenesis family protein
20497	19211	Reverse	Phage Mu protein F like protein
27404	26955	Reverse	Mor transcription activator family protein
34860	32791	Reverse	Mu DNA-binding domain protein
36379	36023	Reverse	DNA-binding transcriptional regulator Nlp
36620	37279	Forward	putative HTH-type transcriptional regulator

Table 46: Functional genes identified for vB_Pae_CF34b.


putative HTHtype transcriptional regulator
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Pae_CF35a genome size was 38.2 kb, the phage was assembled using SPAdes and has a mean coverage of 44. The genome map of this phage is shown in Figure 47. The total number of genes identified was 58 of which 26 were identified with putative functions these can be seen in Table 47. Blastn on viruses only database showed a 1.9 kb / 1.9 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF35a had a 61.4 % G-C content. No tRNA gene was identified however; structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
453	2132	Forward	Phage Terminase
2135	3358	Forward	Phage portal protein
3342	3986	Forward	Caudovirus prohead protease
3983	5197	Forward	Phage capsid family protein
5452	5775	Forward	Phage gp6-like head-tail connector protein
5775	6101	Forward	Phage head-tail joining protein
8473	11748	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
11748	12086	Forward	Phage minor tail protein
12083	12829	Forward	Phage minor tail protein L
12832	13590	Forward	NlpC/P60 family protein
14341	14913	Forward	Bacteriophage lambda tail assembly protein I
14970	18635	Forward	Carbohydrate binding domain protein
20223	19240	Reverse	Transposase DDE domain protein
24931	24701	Reverse	Arc-like DNA binding domain protein
25791	25006	Reverse	BRO family, N-terminal domain
26489	26103	Reverse	Bacterial regulatory proteins, luxR family
27297	26638	Reverse	LexA repressor
27407	27712	Forward	Helix-turn-helix
31668	32483	Forward	DNA replication protein DnaC
32480	33877	Forward	Replicative DNA helicase
34149	34538	Forward	Phage antitermination protein Q
35313	35843	Forward	Phage regulatory protein Rha (Phage_pRha)

35928	36260	Forward	Phage holin family (Lysis protein S)
36257	36874	Forward	Chitinase class I
36874	37386	Forward	Bacteriophage lysis protein
37581	37910	Forward	HNH endonuclease

Table 47: Functional genes identified for vB_Pae_CF35a.

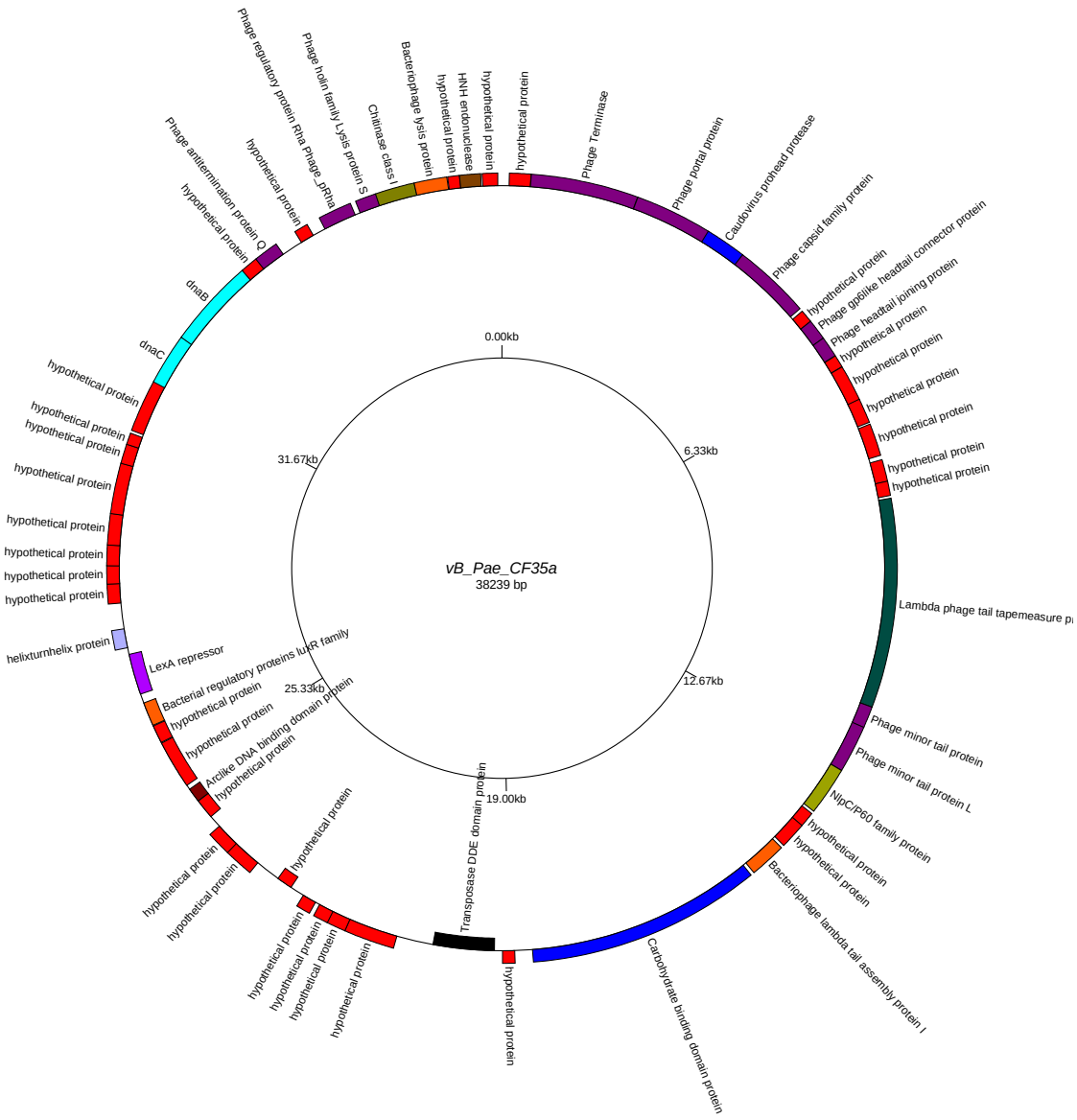


Figure 47: Genome map of vB_Pae_CF35a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The potential incomplete temperate phage vB_Pae_CF35b genome size was 28.2 kb, the phage was assembled using SPAdes and has a mean coverage of 51. The genome map of this phage is shown in Figure 48. The total number of genes identified was 45 of which 7 were identified with putative functions these can be seen in Table 48. Blastn on viruses only database showed a 2.5 kb / 2.6 kb match with *Pseudomonas* phage B3 (GenBank: NC_006548.1). The phage vB_Pae_CF35b had a 63.4 % G-C content. No tRNA gene was identified however; several genes associated with regulatory proteins were identified and an Integrase gene was identified.

Start	End	Direction	Putative functional protein
528	169	Reverse	Mor transcription activator family protein
8047	6263	Reverse	Integrase core domain protein
11056	11445	Forward	Helix-turn-helix domain protein
13317	13946	Forward	Soluble lytic murein transglycosylase precursor
16012	17715	Forward	Terminase-like family protein
19193	20443	Forward	Phage Mu protein F like protein
20440	21012	Forward	Phage virion morphogenesis family protein

Table 48: Functional genes identified for vB_Pae_CF35b.

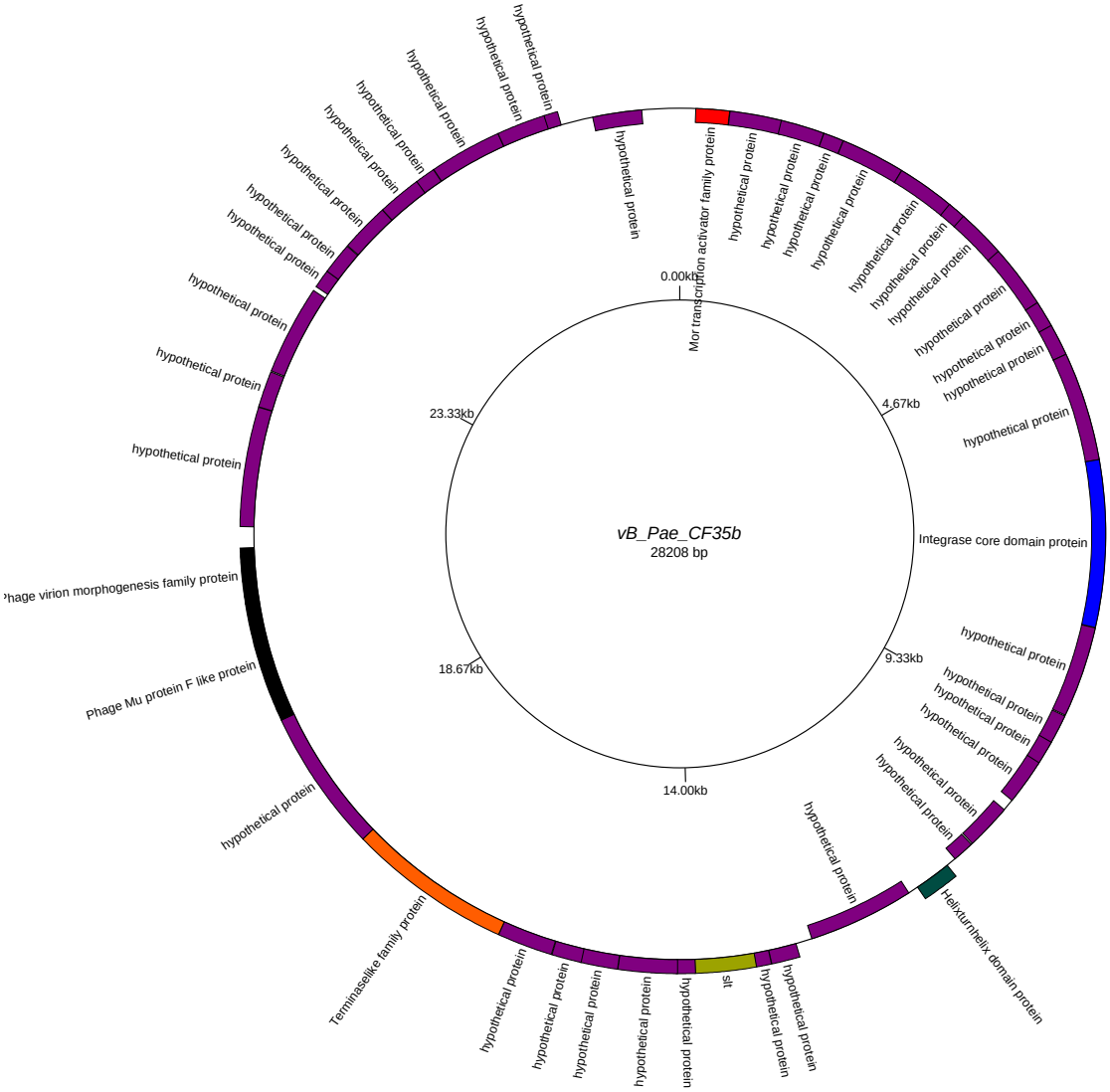


Figure 48: Genome map of vB_Pae_CF35b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF36a genome size was 37.6 kb, the phage was assembled using SPAdes and has a mean coverage of 284. The genome map of this phage is shown in Figure 49. The total number of genes identified was 58 of which 10 were identified with putative functions these can be seen in Table 49. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF36a had a 64.3 % G-C content. No tRNA gene was identified however; Mu-like phage structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
12555	8995	Reverse	Prophage tail length tape measure protein
16495	15587	Reverse	Mu-like prophage major head subunit gpT
18037	16931	Reverse	Mu-like prophage I protein
19523	19056	Reverse	Phage virion morphogenesis family protein
20809	19523	Reverse	Phage Mu protein F like protein
27727	27278	Reverse	Mor transcription activator family protein
31049	30531	Reverse	Bacteriophage Mu Gam like protein
35256	33262	Reverse	Mu DNA-binding domain protein
36314	35958	Reverse	DNA-binding transcriptional regulator Nlp
36484	37197	Forward	HTH-type transcriptional regulator PrtR

Table 49: Functional genes identified for vB_Pae_CF36a.

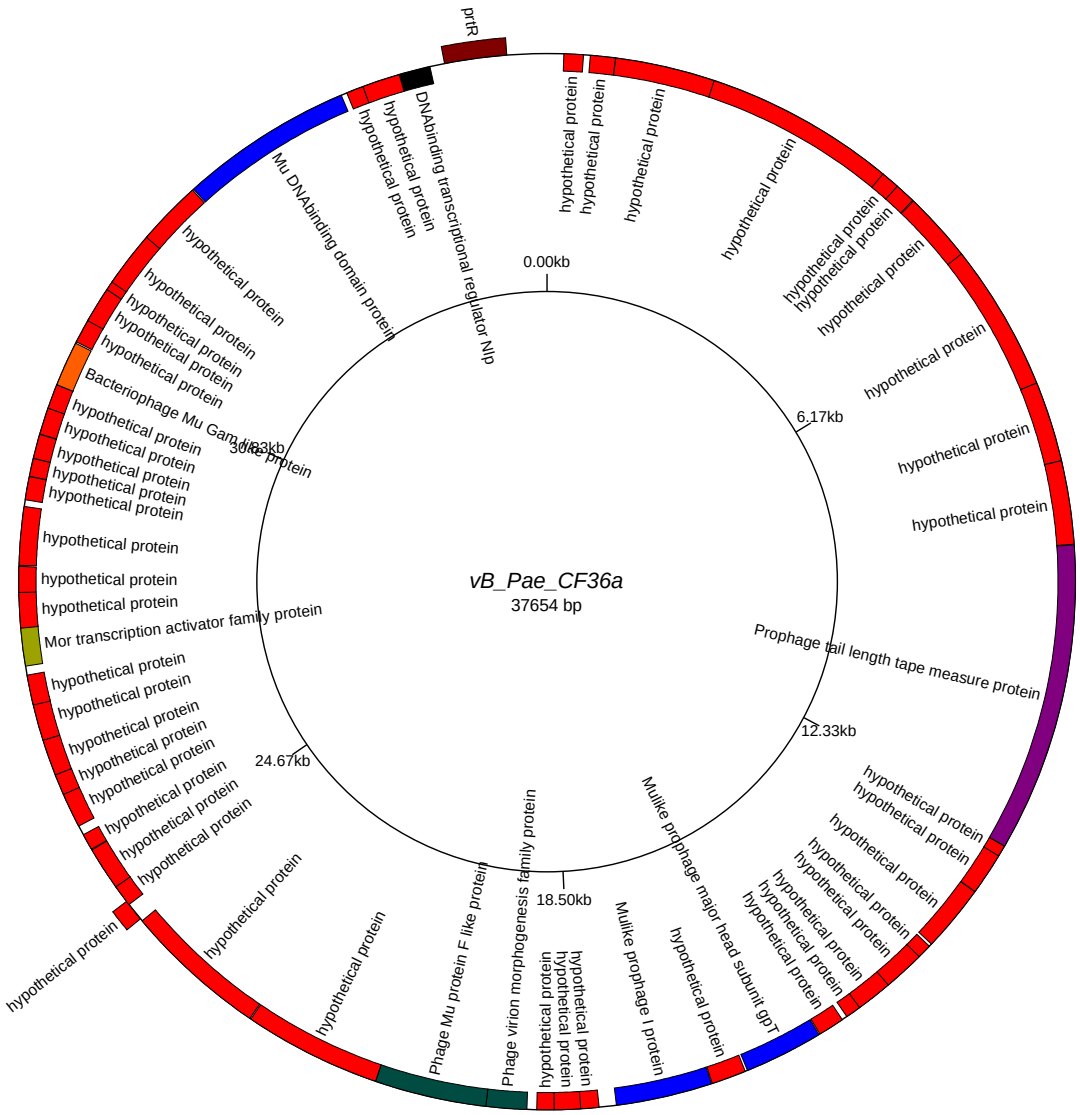
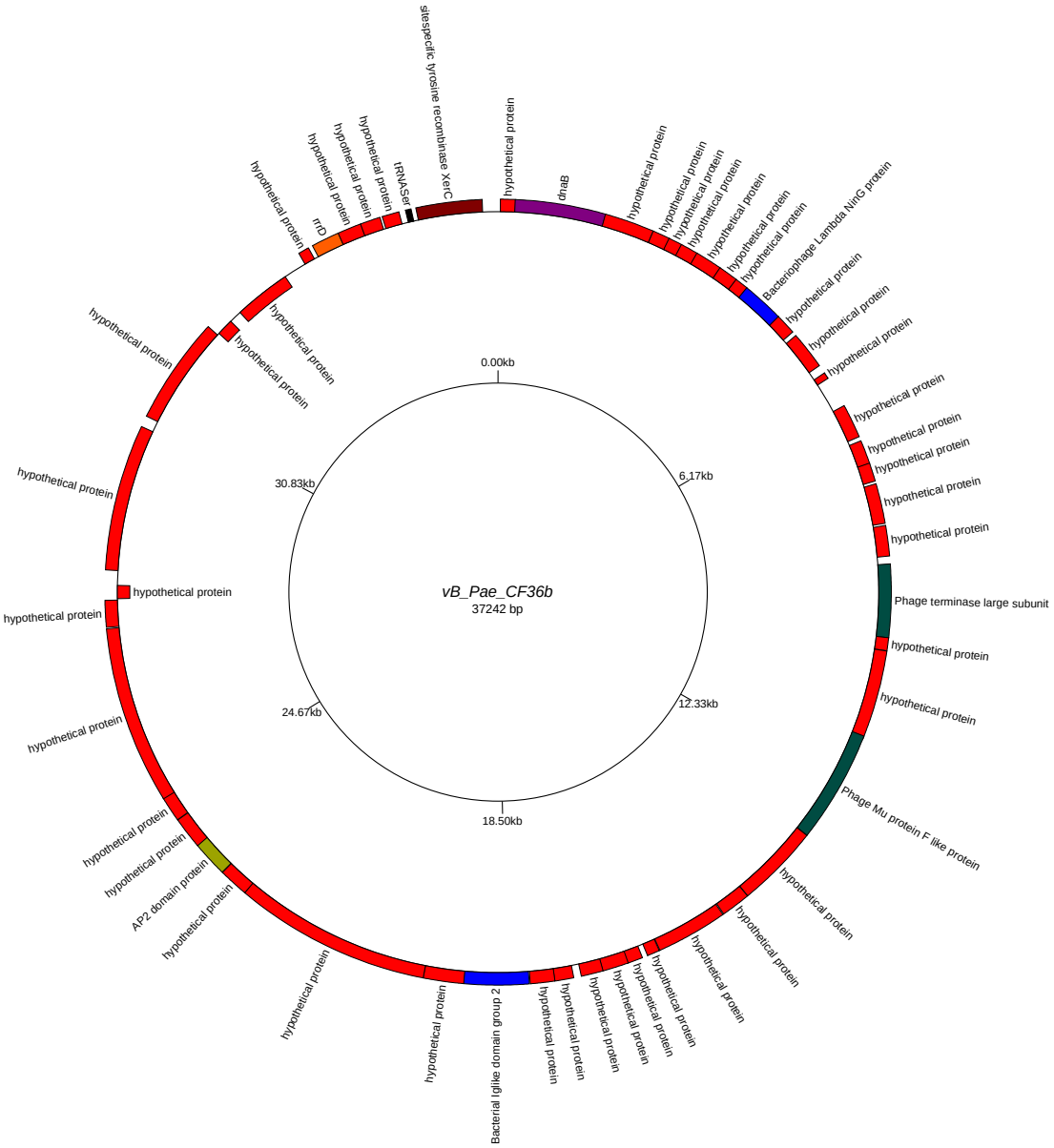


Figure 49: Genome map of vB_Pae_CF36a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF36b genome size was 37.2 kb, the phage was assembled using SPAdes with a mean coverage of 207. The genome map of this phage is shown in Figure 50. The total number of genes identified was 52 of which 11 were identified with putative functions these can be seen in Table 50. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_CF36b had a 59.2 % G-C content. A tRNA gene and an Ig-like domain were identified.

Start	End	Direction	Putative functional protein
259	1644	Forward	Replicative DNA helicase
4058	4705	Forward	Bacteriophage Lambda NinG protein
8883	10001	Forward	Phage terminase large subunit
11534	13264	Forward	Phage Mu protein F like protein
18149	19144	Forward	Bacterial Ig-like domain (group 2)
23230	23790	Forward	AP2 domain protein
30685	32331	Forward	D-glucuronyl C5-epimerase C-terminus
33725	32826	Reverse	BRO family, N-terminal domain
34328	34762	Forward	Lysozyme RrrD
35832	35919	Forward	tRNA-Ser(cga)
35993	37003	Forward	site-specific tyrosine recombinase XerC

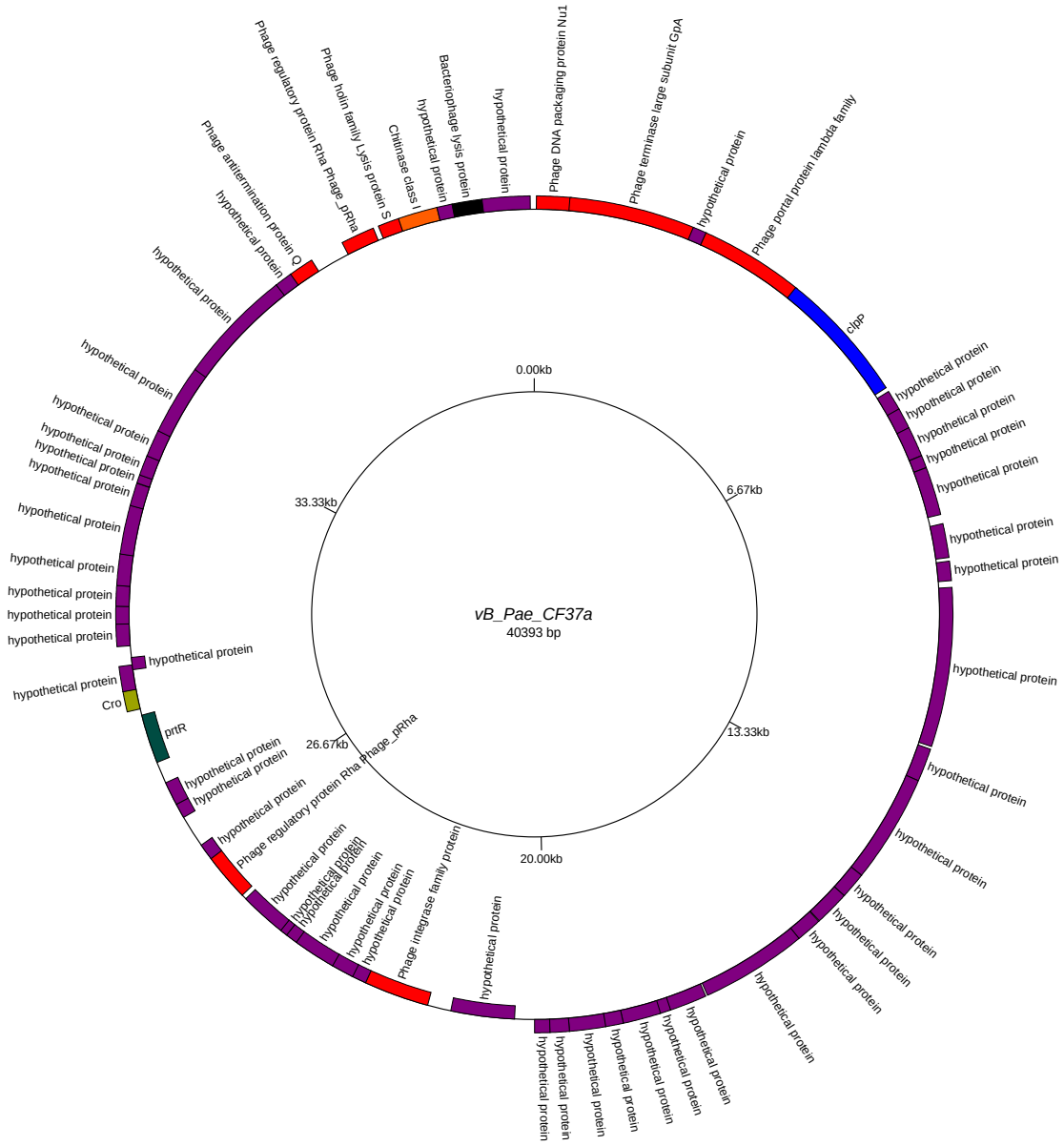
Table 50: Functional genes identified for vB_Pae_CF36b.


vB_Pae_CF36b
37242 bp
0.00kb
6.17kb
12.33kb
18.50kb
24.67kb
30.83kb
site-specific tyrosine recombinase XerC
hypothetical protein
dnaB
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Bacteriophage Lambda NinG protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage terminase large subunit
hypothetical protein
hypothetical protein
Phage Mu protein F like protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Bacterial IgA like domain group 2
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
AP2 domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
tRNASer
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Pae_CF37a genome size was 40.3 kb, the phage was assembled using SPAdes and has a mean coverage of 473. The genome map of this phage is shown in Figure 51. The total number of genes identified was 61 of which 13 were identified with putative functions these can be seen in Table 51. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF37a had a 61.6 % G-C content. No tRNA gene was identified however; a *Cro* gene and phage integrase like gene was identified.

Start	**End**	**Direction**	**Putative functional protein**
32	577	Forward	Phage DNA packaging protein Nu1
549	2513	Forward	Phage terminase large subunit (GpA)
2719	4365	Forward	Phage portal protein, lambda family
4337	6418	Forward	ATP-dependent Clp protease proteolytic subunit
22961	21909	Reverse	Phage integrase family protein
26137	25370	Reverse	Phage regulatory protein Rha (Phage_pRha)
28679	27891	Reverse	HTH-type transcriptional regulator PrtR
28782	29102	Forward	Cro
36398	36787	Forward	Phage antitermination protein Q
37326	37853	Forward	Phage regulatory protein Rha (Phage_pRha)
37938	38270	Forward	Phage holin family (Lysis protein S)
38267	38884	Forward	Chitinase class I
39120	39590	Forward	Bacteriophage lysis protein

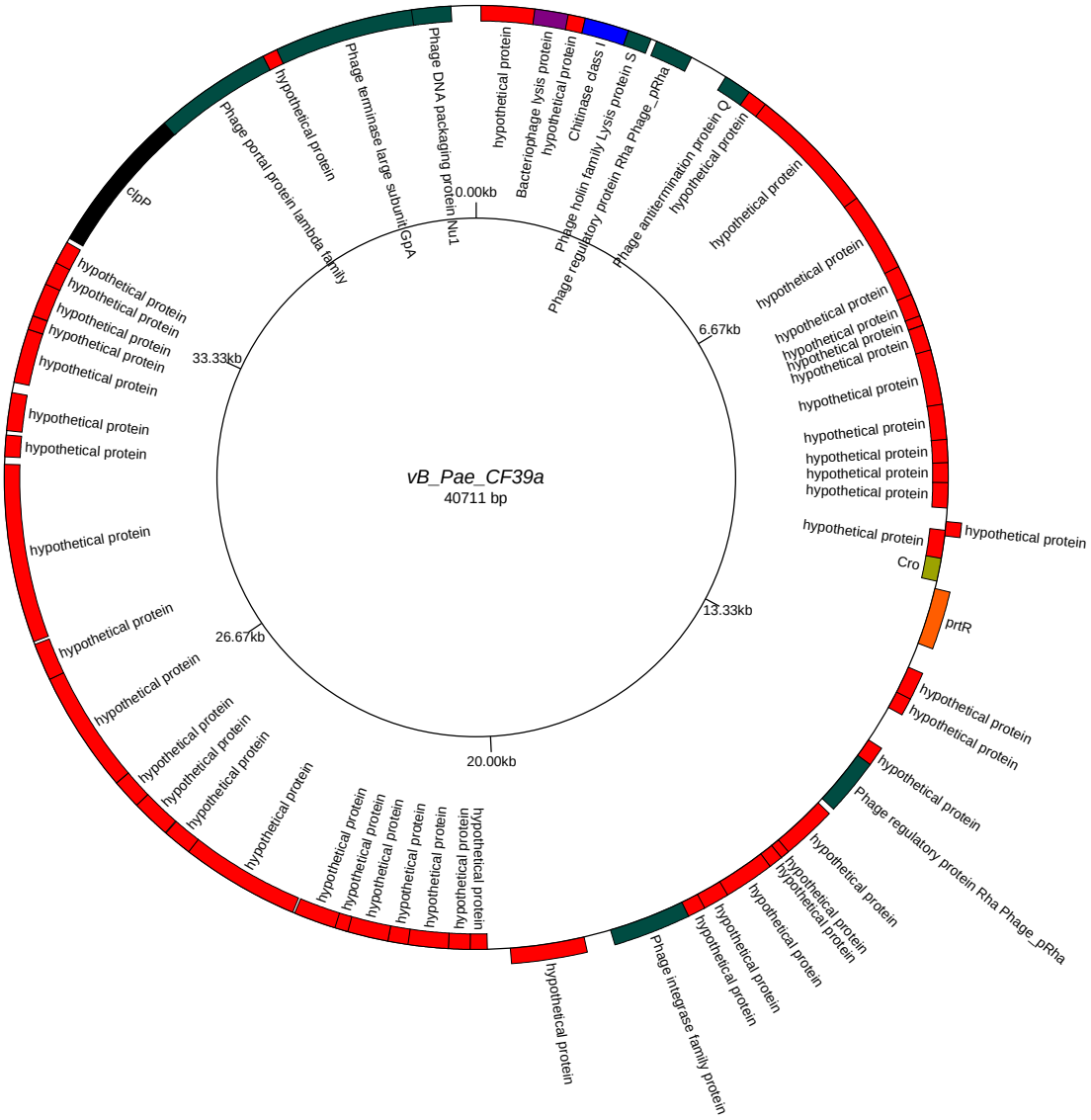
Table 51: Functional genes identified for vB_Pae_CF37a.


vB_Pae_CF37a
40393 bp
0.00kb
6.67kb
13.33kb
20.00kb
26.67kb
33.33kb
Phage DNA packaging protein Nu1
Phage terminase large subunit GpA
hypothetical protein
Phage portal protein lambda family
clpP
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Pae_CF39a genome size was 40.7 kb, the phage was assembled using SPAdes and extended using PriceTI. The genome map of this phage is shown in Figure 52. The total number of genes identified was 61 of which 13 were identified with putative functions these can be seen in Table 52. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF39a had a 61.6 % G-C content. No tRNA gene was identified however; a *Cro* gene and phage integrase like gene was identified.

Start	**End**	**Direction**	**Putative functional protein**
1276	806	Reverse	Bacteriophage lysis protein
2129	1512	Reverse	Chitinase class I
2458	2126	Reverse	Phage holin family (Lysis protein S)
3070	2543	Reverse	Phage regulatory protein Rha (Phage_pRha)
3998	3609	Reverse	Phage antitermination protein Q
11614	11294	Reverse	Cro
11717	12505	Forward	HTH-type transcriptional regulator PrtR
14259	15026	Forward	Phage regulatory protein Rha (Phage_pRha)
17435	18487	Forward	Phage integrase family protein
36059	33978	Reverse	ATP-dependent Clp protease proteolytic subunit
37677	36031	Reverse	Phage portal protein, lambda family
39847	37883	Reverse	Phage terminase large subunit (GpA)
40364	39819	Reverse	Phage DNA packaging protein Nu1

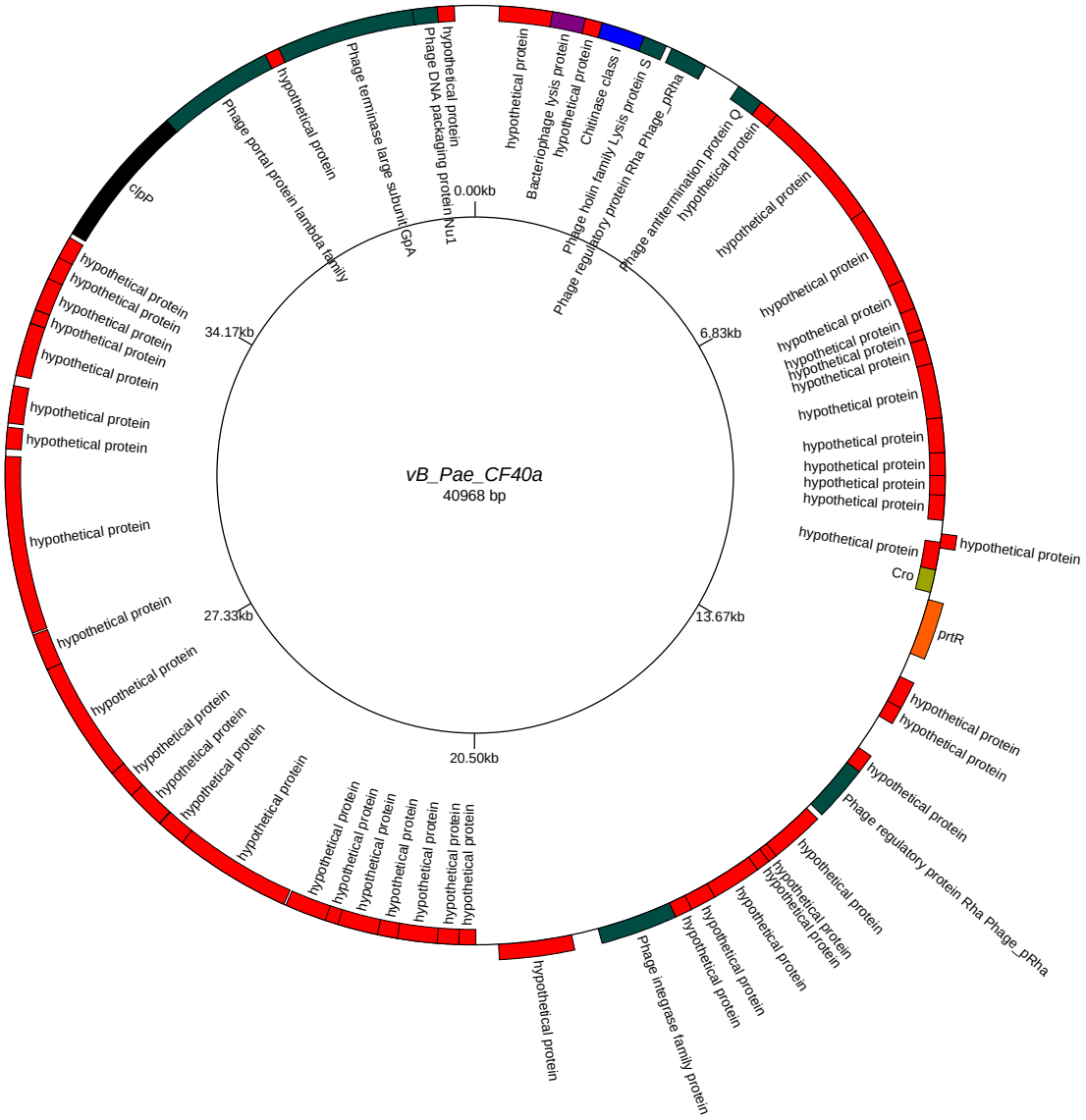
Table 52: Functional genes identified for vB_Pae_CF39a.


vB_Pae_CF39a
40711 bp
0.00kb
6.67kb
13.33kb
20.00kb
26.67kb
33.33kb
Phage DNA packaging protein Nu1
hypothetical protein
Bacteriophage lysis protein
hypothetical protein
Chitinase class I
Phage holin family Lysis protein S
Phage regulatory protein Rha Phage_pRha
Phage antitermination protein Q
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Cro
prtR
hypothetical protein
hypothetical protein
hypothetical protein
Phage regulatory protein Rha Phage_pRha
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage integrase family protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Pae_CF40a genome size was 40.9 kb, the phage was assembled using SPAdes and extended using PriceTI. The genome map of this phage is shown in Figure 53. The total number of genes identified was 62 of which 13 were identified with putative functions these can be seen in Table 53. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF40a had a 61.5 % G-C content. No tRNA gene was identified however; a *Cro* gene and phage integrase like gene was identified.

Start	**End**	**Direction**	Putative functional protein
1549	1079	Reverse	Bacteriophage lysis protein
2402	1785	Reverse	Chitinase class I
2731	2399	Reverse	Phage holin family (Lysis protein S)
3343	2816	Reverse	Phage regulatory protein Rha (Phage_pRha)
4271	3882	Reverse	Phage antitermination protein Q
11887	11567	Reverse	Cro
11990	12778	Forward	HTH-type transcriptional regulator PrtR
14532	15299	Forward	Phage regulatory protein Rha (Phage_pRha)
17708	18760	Forward	Phage integrase family protein
36332	34251	Reverse	ATP-dependent Clp protease proteolytic subunit
37950	36304	Reverse	Phage portal protein, lambda family
40120	38156	Reverse	Phage terminase large subunit (GpA)
40490	40092	Reverse	Phage DNA packaging protein Nu1

Table 53: Functional genes identified for vB_Pae_CF40a.


vB_Pae_CF40a
40968 bp
0.00kb
6.83kb
13.67kb
20.50kb
27.33kb
34.17kb
hypothetical protein
Phage DNA packaging protein Nu1
hypothetical protein
Phage terminase large subunit GpA
hypothetical protein
Phage portal protein lambda family
ClpP
hypothetical protein
hypothetical protein
hypothetical protein

The potential partial temperate phage vB_Pae_CF41a genome size was 37.3 kb, the phage was assembled using SPAdes and has a mean coverage of 55. The genome map of this phage is shown in Figure 54. The total number of genes identified was 54 of which 8 were identified with putative functions these can be seen in Table 54. Blastn on viruses only database showed a 7 kb / 7.5 kb match with *Pseudomonas* phage B3 (GenBank: NC_006548.1). The phage vB_Pae_CF41a had a 63.6 % G-C content. No tRNA gene was identified however; several genes associated with regulatory proteins were identified and an Integrase gene was identified.

Start	End	Direction	Putative functional protein
8314	7742	Reverse	Phage virion morphogenesis family protein
9561	8311	Reverse	Phage Mu protein F like protein
12742	11039	Reverse	Terminase-like family protein
15437	14808	Reverse	Soluble lytic murein transglycosylase precursor
17698	17309	Reverse	Helix-turn-helix domain protein
20707	22491	Forward	Integrase core domain protein
28226	28585	Forward	Mor transcription activator family protein
28775	30106	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)

Table 54: Functional genes identified for vB_Pae_CF41a.

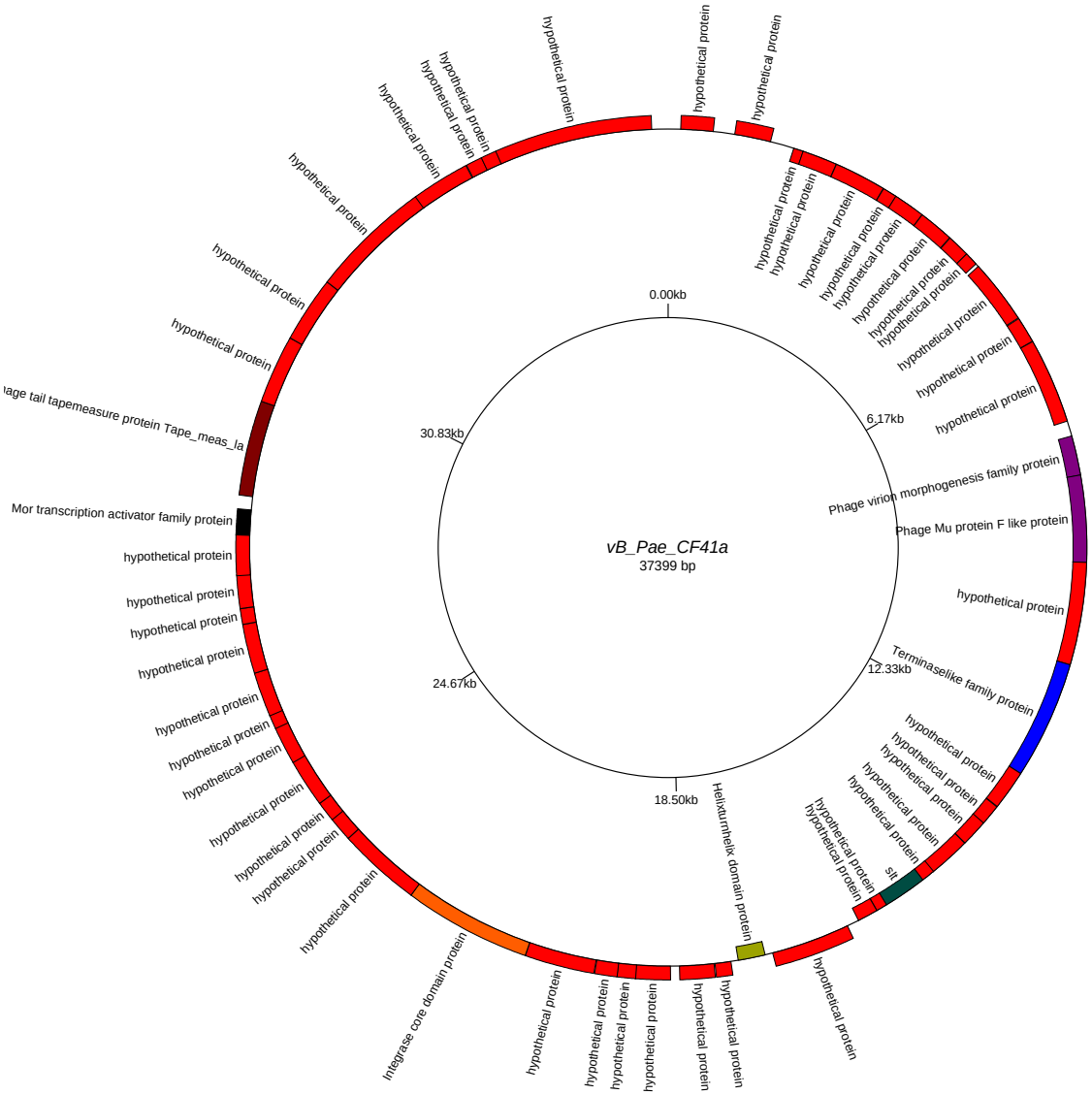


Figure 54: Genome map of vB_Pae_CF41a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF41b genome size was 37.6 kb, the phage was assembled using SPAdes and has a mean coverage of 187. The genome map of this phage is shown in Figure 55. The total number of genes identified was 58 of which 10 were identified with putative functions these can be seen in Table 55. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF41b had a 64.3 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
1123	410	Reverse	HTH-type transcriptional regulator PrtR
1293	1649	Forward	DNA-binding transcriptional regulator Nlp
2351	4345	Forward	Mu DNA-binding domain protein
6558	7076	Forward	Bacteriophage Mu Gam like protein
9880	10329	Forward	Mor transcription activator family protein
16798	18084	Forward	Phage Mu protein F like protein
18084	18551	Forward	Phage virion morphogenesis family protein
19570	20676	Forward	Mu-like prophage I protein
21112	22020	Forward	Mu-like prophage major head subunit gpT
25052	28612	Forward	Prophage tail length tape measure protein

Table 55: Functional genes identified for vB_Pae_CF41b.

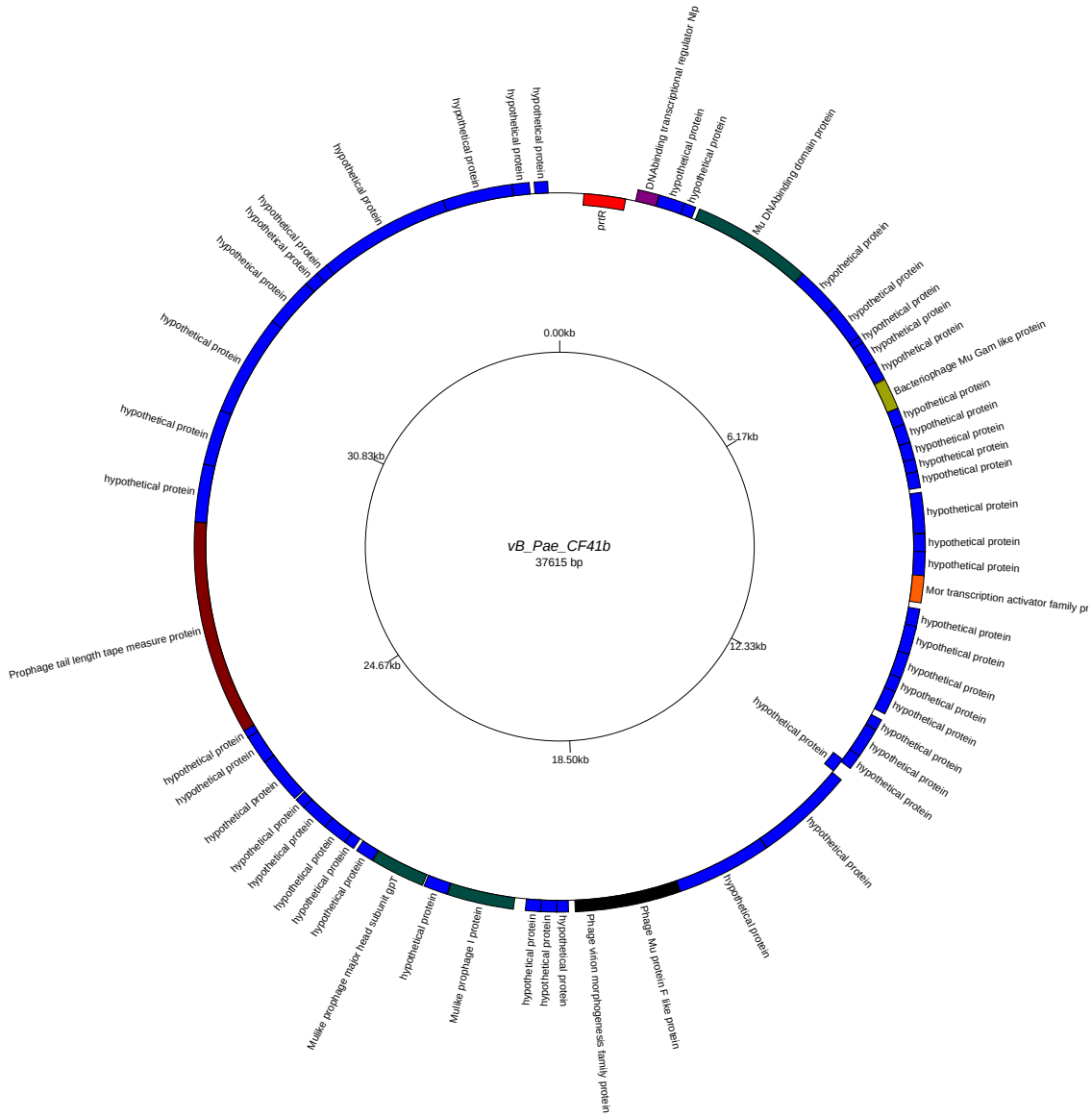
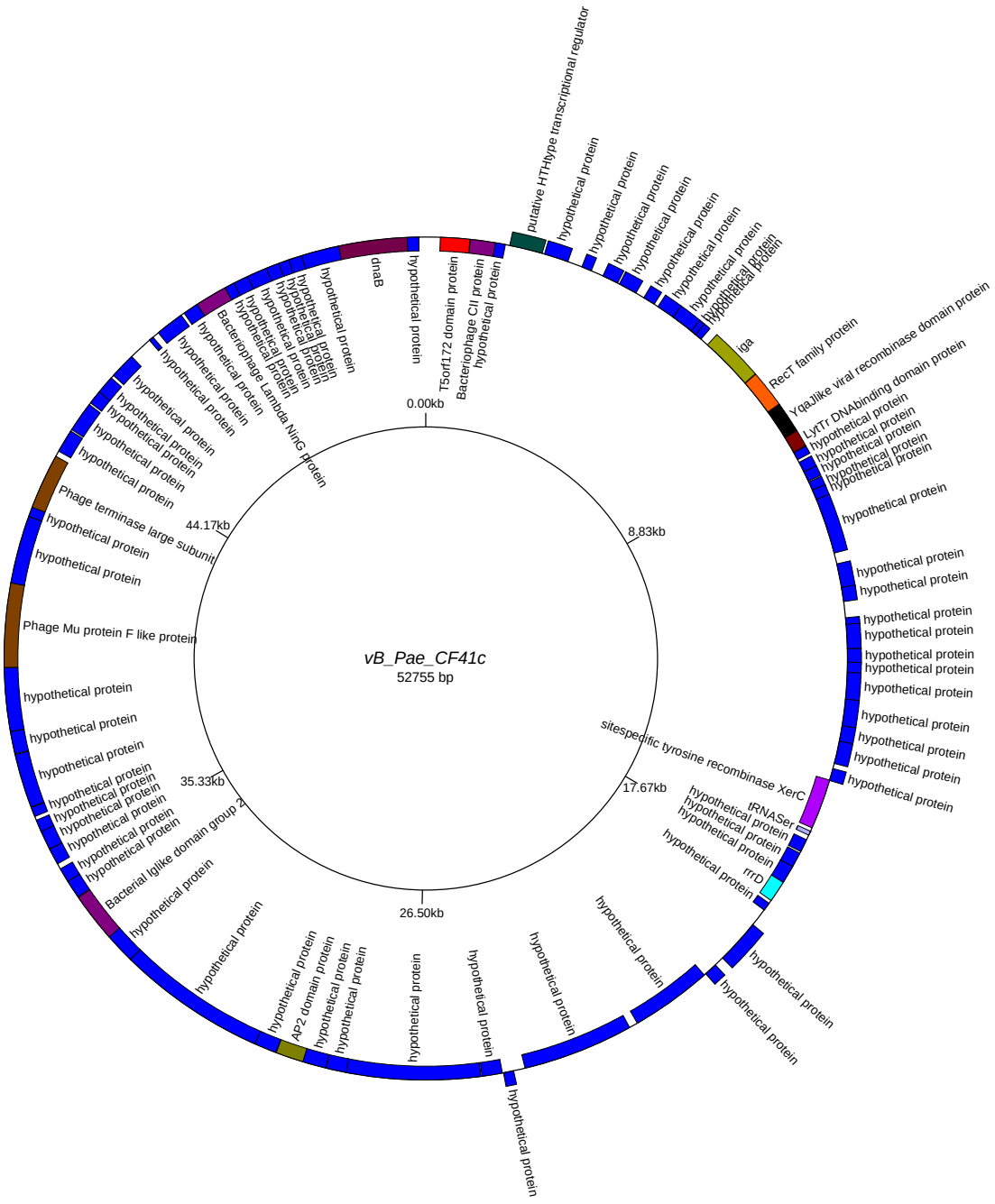


Figure 55: Genome map of vB_Pae_CF41b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF41c genome size was 52.7 kb, the phage was assembled using SPAdes with a mean coverage of 379. The genome map of this phage is shown in Figure 56. The total number of genes identified was 86 of which 18 were identified with putative functions these can be seen in Table 56. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_CF41c had a 58.9 % G-C content. A tRNA gene was identified along with genes associated with structural, lysogenic and regulatory proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
886	290	Reverse	T5orf172 domain protein
1393	890	Reverse	Bacteriophage CII protein
1683	2348	Forward	putative HTH-type transcriptional regulator
6142	7206	Forward	IgA-specific serine endopeptidase autotransporter precursor
7214	7960	Forward	RecT family protein
7944	8564	Forward	YqaJ-like viral recombinase domain protein
8561	8896	Forward	LytTr DNA-binding domain protein
16658	15648	Reverse	site-specific tyrosine recombinase XerC
16819	16732	Reverse	tRNA-Ser(cga)
18323	17889	Reverse	Lysozyme RrrD
18926	19825	Forward	BRO family, N-terminal domain
21966	20320	Reverse	D-glucuronyl C5-epimerase C-terminus
29421	28861	Reverse	AP2 domain protein
34502	33507	Reverse	Bacterial Ig-like domain (group 2)
41117	39387	Reverse	Phage Mu protein F like protein
43768	42650	Reverse	Phage terminase large subunit
48593	47946	Reverse	Bacteriophage Lambda NinG protein
52392	51007	Reverse	Replicative DNA helicase

Table 56: Functional genes identified for vB_Pae_CF41c.


vB_Pae_CF41c
52755 bp
0.00kb
8.83kb
17.67kb
26.50kb
35.33kb
44.17kb
putative HTHtype transcriptional regulator
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
tga
RecT family protein
YqaJlike viral recombinase domain protein
LytTr DNAbinding domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Pae_CF42a genome size was 37.6 kb, the phage was assembled using SPAdes and has a mean coverage of 51. The genome map of this phage is shown in Figure 57. The total number of genes identified was 57 of which 26 were identified with putative functions these can be seen in Table 57. Blastn on viruses only database showed a 1.9 kb / 1.9 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF42a had a 61.4 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
101	1483	Forward	Phage Terminase
1486	2709	Forward	Phage portal protein
2693	3337	Forward	Caudovirus prohead protease
3334	4548	Forward	Phage capsid family protein
4803	5126	Forward	Phage gp6-like head-tail connector protein
5126	5452	Forward	Phage head-tail joining protein
7824	11099	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
11099	11437	Forward	Phage minor tail protein
11434	12180	Forward	Phage minor tail protein L
12183	12941	Forward	NlpC/P60 family protein
13692	14264	Forward	Bacteriophage lambda tail assembly protein I
14321	17986	Forward	Carbohydrate binding domain protein
19574	18591	Reverse	Transposase DDE domain protein
24282	24052	Reverse	Arc-like DNA binding domain protein
25142	24357	Reverse	BRO family, N-terminal domain
25840	25454	Reverse	Bacterial regulatory proteins, luxR family
26648	25989	Reverse	LexA repressor
26758	27063	Forward	Helix-turn-helix
31019	31834	Forward	DNA replication protein DnaC
31831	33228	Forward	Replicative DNA helicase
33500	33889	Forward	Phage antitermination protein Q

34664	35194	Forward	Phage regulatory protein Rha (Phage_pRha)
35279	35611	Forward	Phage holin family (Lysis protein S)
35608	36225	Forward	Chitinase class I
36225	36737	Forward	Bacteriophage lysis protein
36932	37261	Forward	HNH endonuclease

Table 57: Functional genes identified for vB_Pae_CF42a.

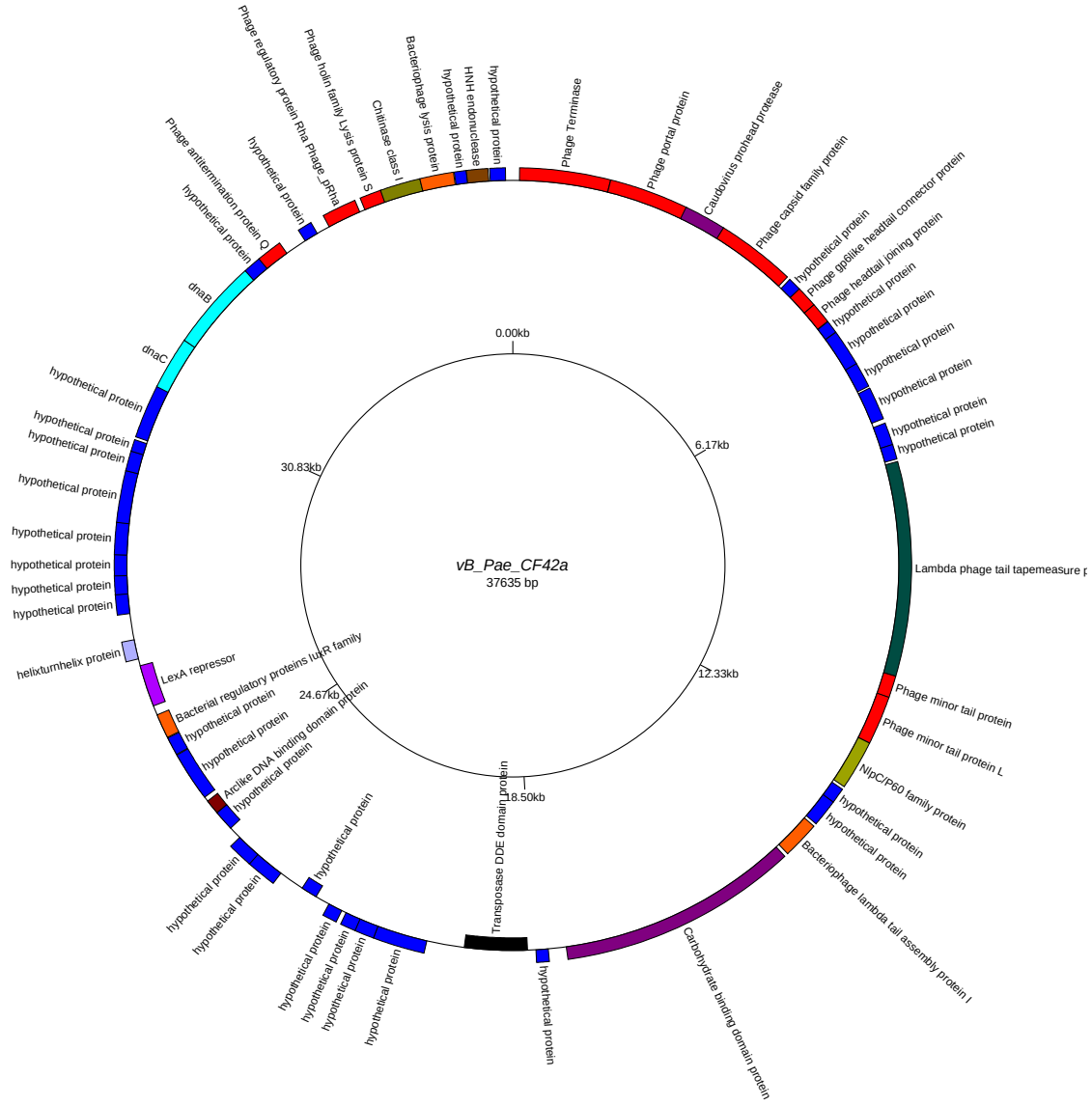


Figure 57: Genome map of vB_Pae_CF42a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF42b genome size was 38.4 kb, the phage was assembled using SPAdes and has a mean coverage of 258. The genome map of this phage is shown in Figure 58. The total number of genes identified was 59 of which 10 were identified with putative functions these can be seen in Table 58. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF42b had a 64.2 % G-C content. No tRNA gene was identified however; Mu-like phage structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
12580	9020	Reverse	Prophage tail length tape measure protein
16520	15612	Reverse	Mu-like prophage major head subunit gpT
18062	16956	Reverse	Mu-like prophage I protein
19548	19081	Reverse	Phage virion morphogenesis family protein
20834	19548	Reverse	Phage Mu protein F like protein
27752	27303	Reverse	Mor transcription activator family protein
31074	30556	Reverse	Bacteriophage Mu Gam like protein
35281	33287	Reverse	Mu DNA-binding domain protein
36339	35983	Reverse	DNA-binding transcriptional regulator Nlp
36509	37222	Forward	HTH-type transcriptional regulator PrtR

Table 58: Functional genes identified for vB_Pae_CF42b.

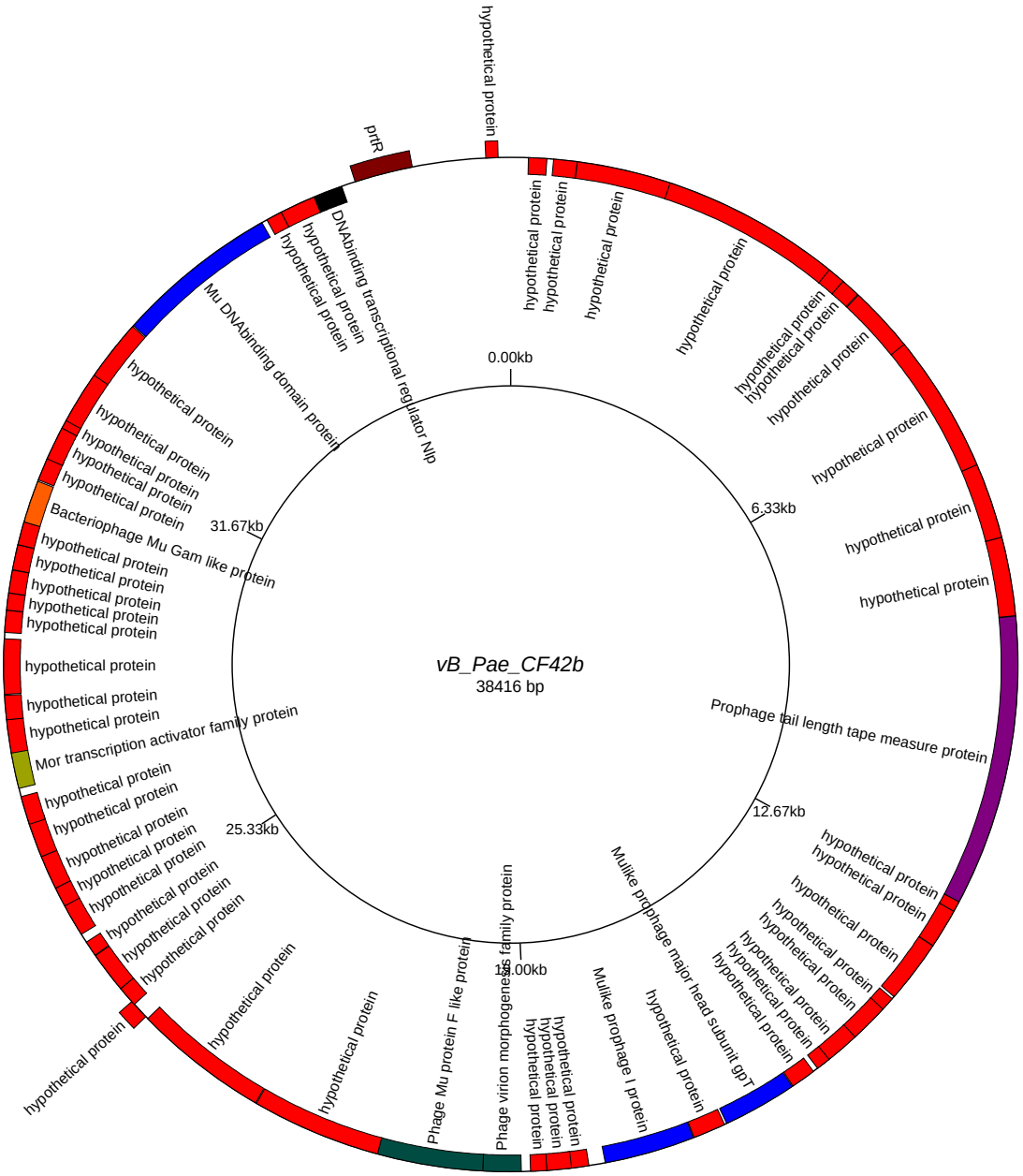


Figure 58: Genome map of vB_Pae_CF42b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF43a genome size was 40.5 kb, the phage was assembled using SPAdes and extended with PriceTI. The genome map of this phage is shown in Figure 59. The total number of genes identified was 63 of which 12 were identified with putative functions these can be seen in Table 59. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_CF43a had a 61.5 % G-C content. No tRNA gene was identified however; a *Cro* gene and phage integrase like gene was identified.

Start	**End**	**Direction**	**Putative functional protein**
678	2642	Forward	Phage terminase large subunit (GpA)
2848	4494	Forward	Phage portal protein, lambda family
4466	6547	Forward	ATP-dependent Clp protease proteolytic subunit
23090	22038	Reverse	Phage integrase family protein
26266	25499	Reverse	Phage regulatory protein Rha (Phage_pRha)
28808	28020	Reverse	HTH-type transcriptional regulator PrtR
28911	29231	Forward	Cro
36528	36917	Forward	Phage antitermination protein Q
37456	37983	Forward	Phage regulatory protein Rha (Phage_pRha)
38068	38400	Forward	Phage holin family (Lysis protein S)
38397	39014	Forward	Chitinase class I
39250	39720	Forward	Bacteriophage lysis protein

Table 59: Functional genes identified for vB_Pae_CF43a.

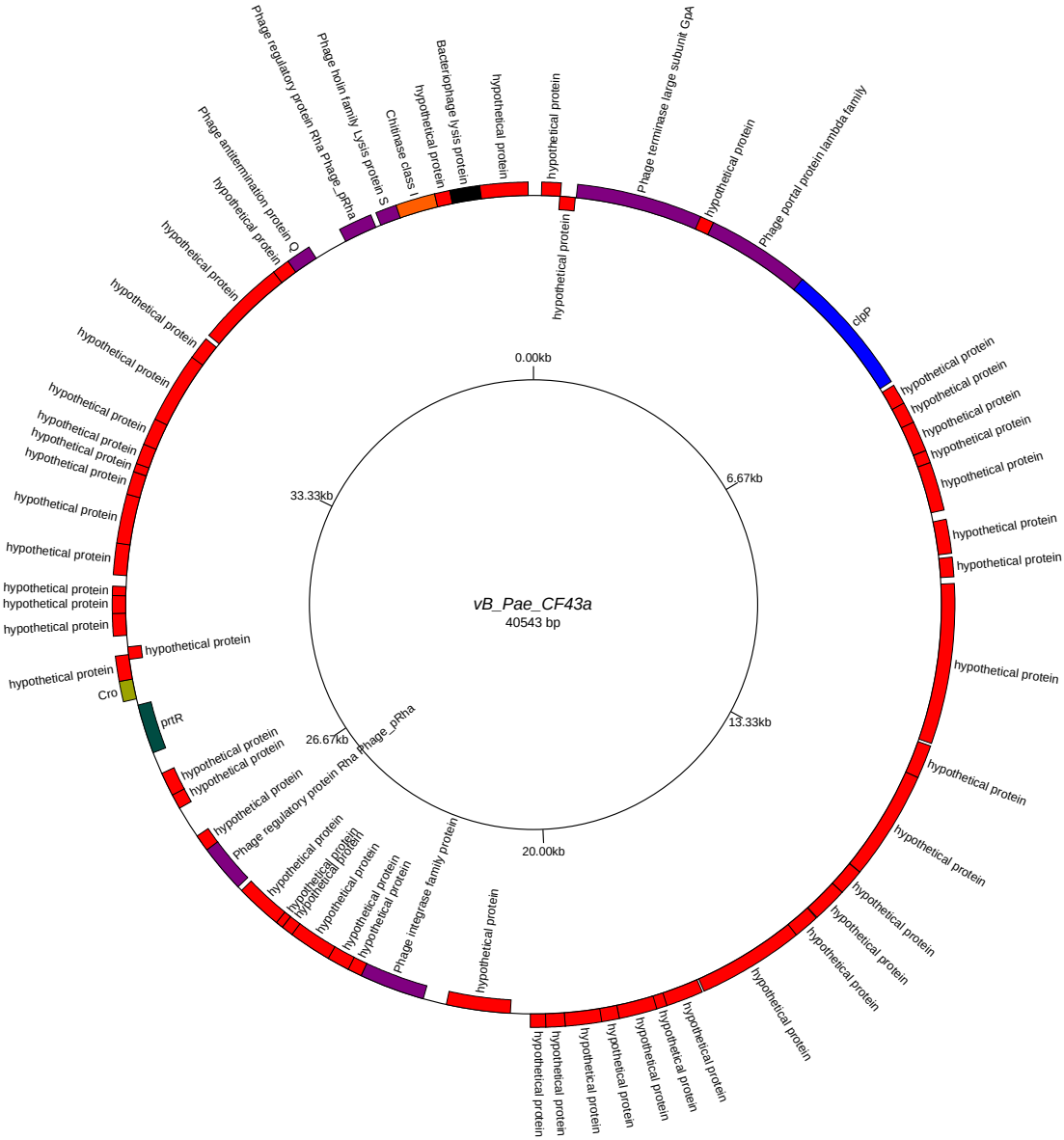


Figure 59: Genome map of vB_Pae_CF43a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF44a genome size was 52.6 kb, the phage was assembled using SPAdes with a mean coverage of 204. The genome map of this phage is shown in Figure 60. The total number of genes identified was 86 of which 18 were identified with putative functions these can be seen in Table 60. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_CF44a had a 58.9 % G-C content. A tRNA gene was identified along with genes associated with structural, lysogenic and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
849	253	Reverse	T5orf172 domain protein
1356	853	Reverse	Bacteriophage CII protein
1646	2311	Forward	putative HTH-type transcriptional regulator
6105	7169	Forward	IgA-specific serine endopeptidase autotransporter precursor
7177	7923	Forward	RecT family protein
7907	8527	Forward	YqaJ-like viral recombinase domain protein
8524	8859	Forward	LytTr DNA-binding domain protein
16621	15611	Reverse	site-specific tyrosine recombinase XerC
16782	16695	Reverse	tRNA-Ser(cga)
18286	17852	Reverse	Lysozyme RrrD
18889	19788	Forward	BRO family, N-terminal domain
21929	20283	Reverse	D-glucuronyl C5-epimerase C-terminus
29384	28824	Reverse	AP2 domain protein
34465	33470	Reverse	Bacterial Ig-like domain (group 2)
41080	39350	Reverse	Phage Mu protein F like protein
43731	42613	Reverse	Phage terminase large subunit
48556	47909	Reverse	Bacteriophage Lambda NinG protein
52355	50970	Reverse	Replicative DNA helicase

Table 60: Functional genes identified for vB_Pae_CF44a.

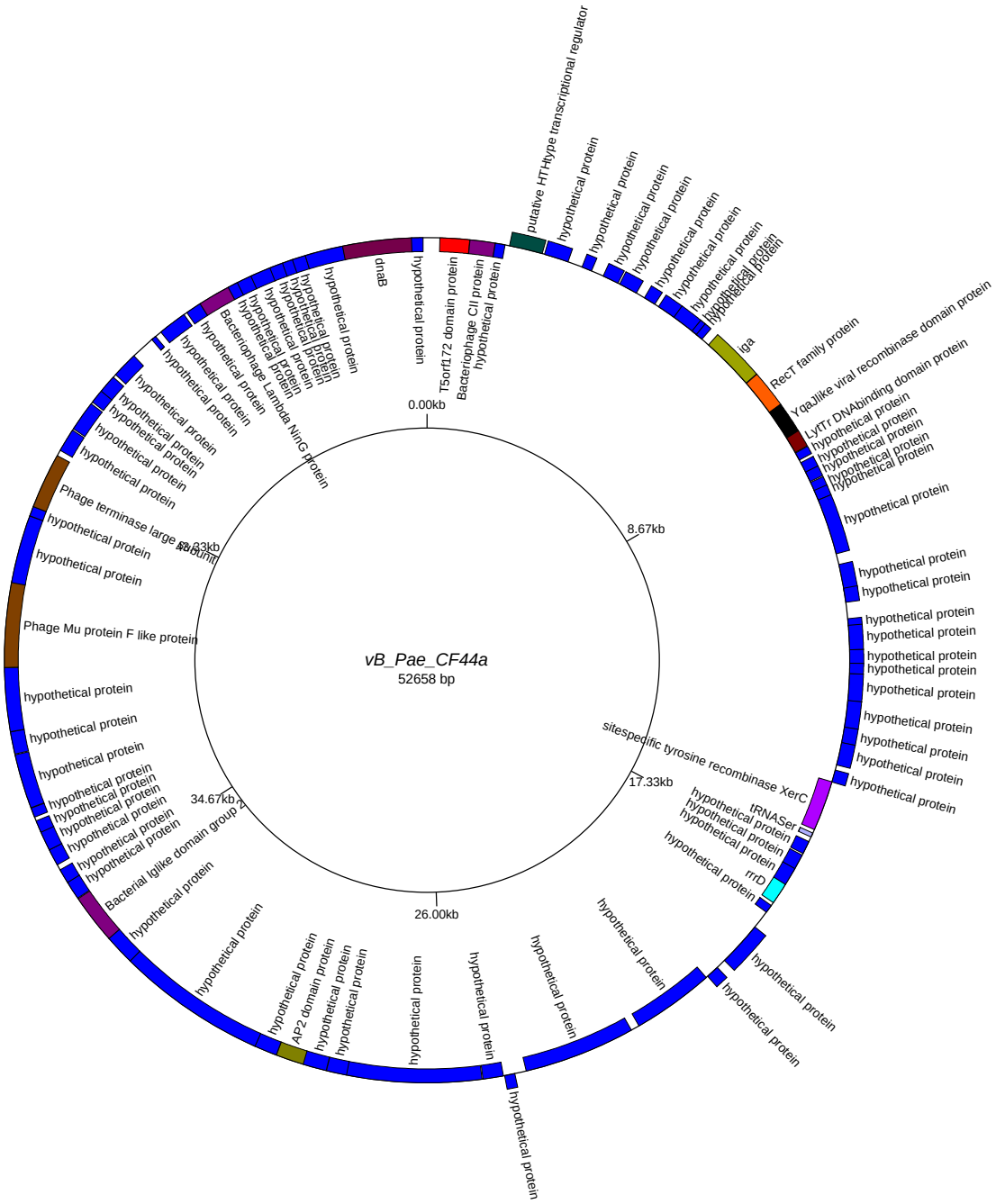
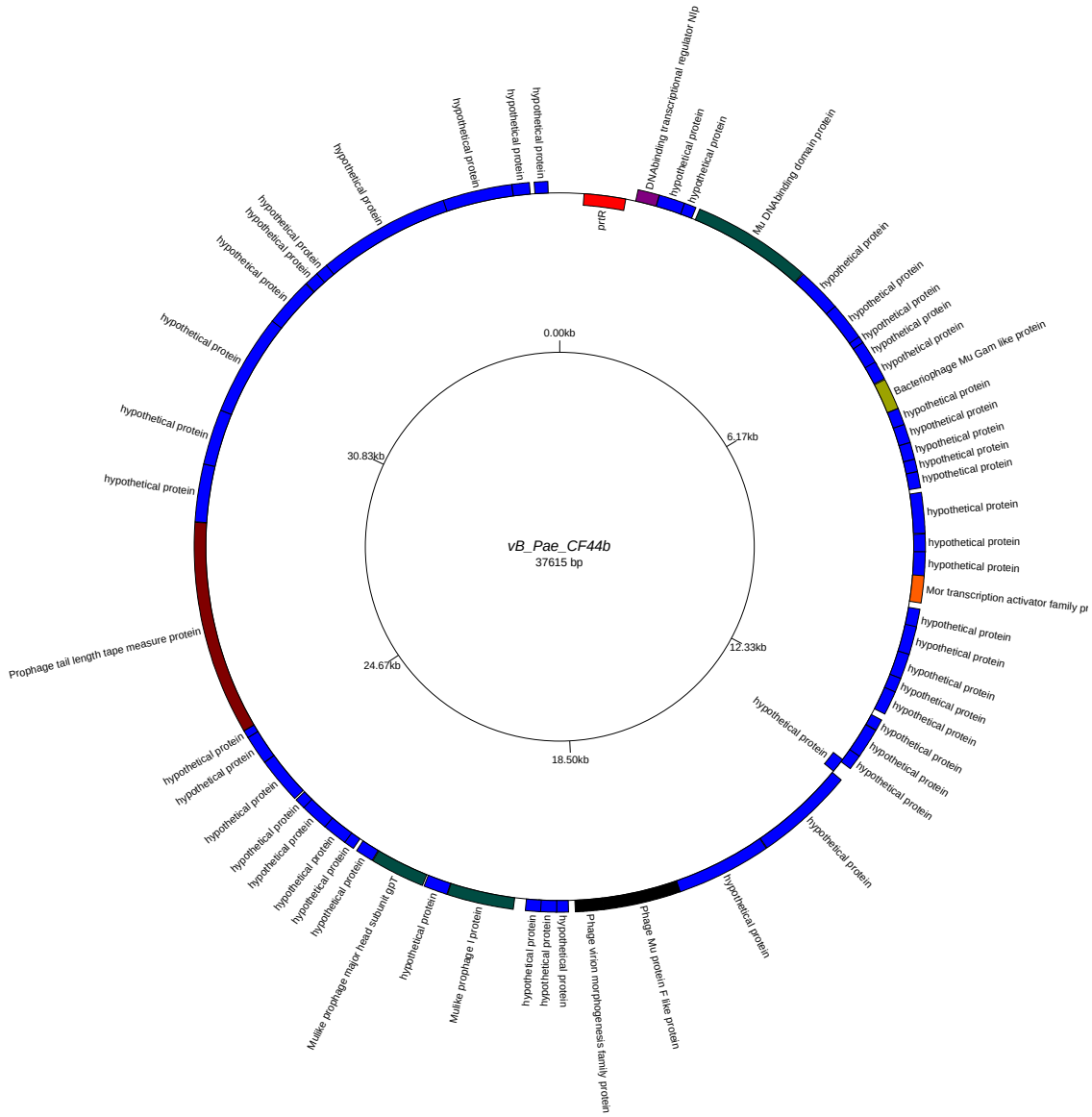


Figure 60: Genome map of vB_Pae_CF44a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_CF44b genome size was 37.6 kb, the phage was assembled using SPAdes and has a mean coverage of 252. The genome map of this phage is shown in Figure 61. The total number of genes identified was 58 of which 10 were identified with putative functions these can be seen in Table 61. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_CF44b had a 64.3 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
1124	411	Reverse	HTH-type transcriptional regulator PrtR
1294	1650	Forward	DNA-binding transcriptional regulator Nlp
2352	4346	Forward	Mu DNA-binding domain protein
6559	7077	Forward	Bacteriophage Mu Gam like protein
9881	10330	Forward	Mor transcription activator family protein
16799	18085	Forward	Phage Mu protein F like protein
18085	18552	Forward	Phage virion morphogenesis family protein
19571	20677	Forward	Mu-like prophage I protein
21113	22021	Forward	Mu-like prophage major head subunit gpT
25053	28613	Forward	Prophage tail length tape measure protein

Table 61: Functional genes identified for vB_Pae_CF44b.


vB_Pae_CF44b
37615 bp
0.00kb
6.17kb
12.33kb
18.50kb
24.67kb
30.83kb
orfR
DNAbinding transcriptional regulator Nip
hypothetical protein
hypothetical protein
Mu DNAbinding domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Bacteriophage Mu Gam like protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Mor transcription activator family pr
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage Mu protein F like protein
Phage virion morphogenesis family protein
hypothetical protein
hypothetical protein
hypothetical protein
Mulike prophage I protein
hypothetical protein
Mulike prophage major head subunit gpT
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Prophage tail length tape measure protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The potential partial temperate phage vB_Pae_BR46a genome size was 61.6 kb, the phage was assembled using SPAdes and extended using PriceTI. The genome map of this phage is shown in Figure 62. The total number of genes identified was 63 of which 13 were identified with putative functions these can be seen in Table 62. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66 (GenBank: KC262634.1). The phage vB_Pae_BR46a had a 63.7 % G-C content. No tRNA gene was identified however; a *CPS-53* integrase like gene, *Cro* gene was identified.

Start	End	Direction	Putative functional protein
654	1151	Forward	Single-stranded DNA-binding protein
1178	2398	Forward	Recombination-associated protein RdgC
2395	2604	Forward	LexA repressor
2749	5379	Forward	DNA methylase
5376	7184	Forward	C-5 cytosine-specific DNA methylase
9713	9946	Forward	Response regulator inhibitor for tor operon
11182	9947	Reverse	Putative prophage CPS-53 integrase
14787	14248	Reverse	Phage lysozyme
50010	48727	Reverse	Phage terminase large subunit
51808	51194	Reverse	Bacteriophage Lambda NinG protein
54414	54214	Reverse	Cro
54522	55319	Forward	HTH-type transcriptional regulator PrtR
56738	57109	Forward	Carbon storage regulator homolog

Table 62: Functional genes identified for vB_Pae_BR46a.

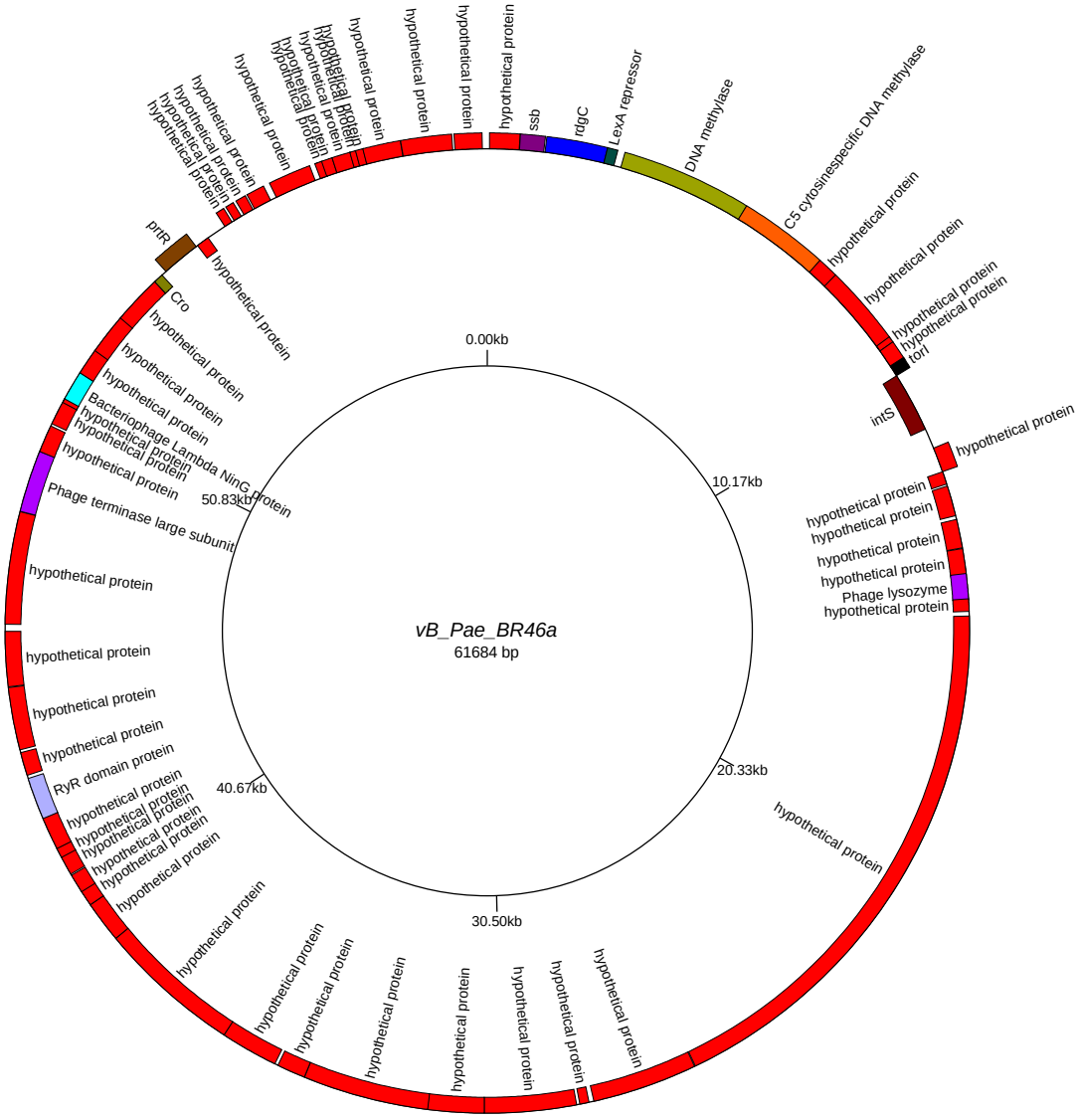


Figure 62: Genome map of vB_Pae_BR46a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR47a genome size was 38.3 kb, the phage was assembled using SPAdes and has a mean coverage of 64. The genome map of this phage is shown in Figure 63. The total number of genes identified was 58 of which 26 were identified with putative functions these can be seen in Table 63. Blastn on viruses only database showed a 1.9 kb / 1.9 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR47a had a 61.4 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
688	359	Reverse	HNH endonuclease
1395	883	Reverse	Bacteriophage lysis protein
2012	1395	Reverse	Chitinase class I
2341	2009	Reverse	Phage holin family (Lysis protein S)
2956	2426	Reverse	Phage regulatory protein Rha (Phage_pRha)
4120	3731	Reverse	Phage antitermination protein Q
5789	4392	Reverse	Replicative DNA helicase
6601	5786	Reverse	DNA replication protein DnaC
10862	10557	Reverse	Helix-turn-helix
10972	11631	Forward	LexA repressor
11780	12166	Forward	Bacterial regulatory proteins, luxR family
12478	13263	Forward	BRO family, N-terminal domain
13338	13568	Forward	Arc-like DNA binding domain protein
18046	19029	Forward	Transposase DDE domain protein
23299	19634	Reverse	Carbohydrate binding domain protein
23928	23356	Reverse	Bacteriophage lambda tail assembly protein I
25437	24679	Reverse	NlpC/P60 family protein
26186	25440	Reverse	Phage minor tail protein L
26521	26183	Reverse	Phage minor tail protein
29796	26521	Reverse	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
32494	32168	Reverse	Phage head-tail joining protein

32817	32494	Reverse	Phage gp6-like head-tail connector protein
34286	33072	Reverse	Phage capsid family protein
34927	34283	Reverse	Caudovirus prohead protease
36134	34911	Reverse	Phage portal protein
37816	36137	Reverse	Phage Terminase

Table 63: Functional genes identified for vB_Pae_BR47a.

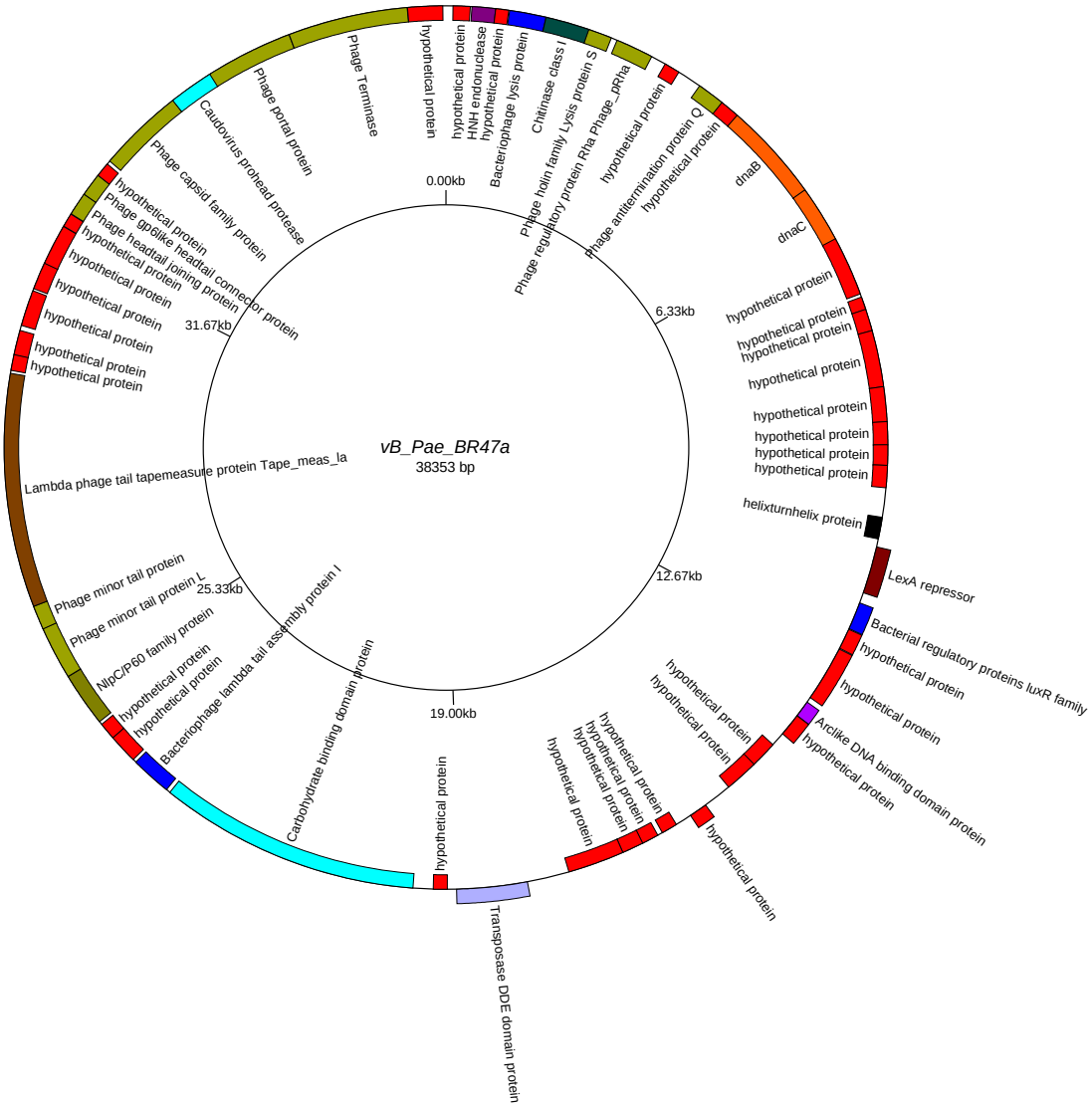


Figure 63: Genome map of vB_Pae_BR47a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR47b genome size was 31 kb, the phage was assembled using SPAdes and has a mean coverage of 218. The genome map of this phage is shown in Figure 64. The total number of genes identified was 48 of which 7 were identified with putative functions these can be seen in Table 64. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_BR47b had a 64.5 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
79	471	Forward	Bacteriophage Mu Gam like protein
3275	3724	Forward	Mor transcription activator family protein
10193	11479	Forward	Phage Mu protein F like protein
11479	11946	Forward	Phage virion morphogenesis family protein
12965	14071	Forward	Mu-like prophage I protein
14507	15415	Forward	Mu-like prophage major head subunit gpT
18447	22007	Forward	Prophage tail length tape measure protein

Table 64: Functional genes identified for vB_Pae_BR47b.

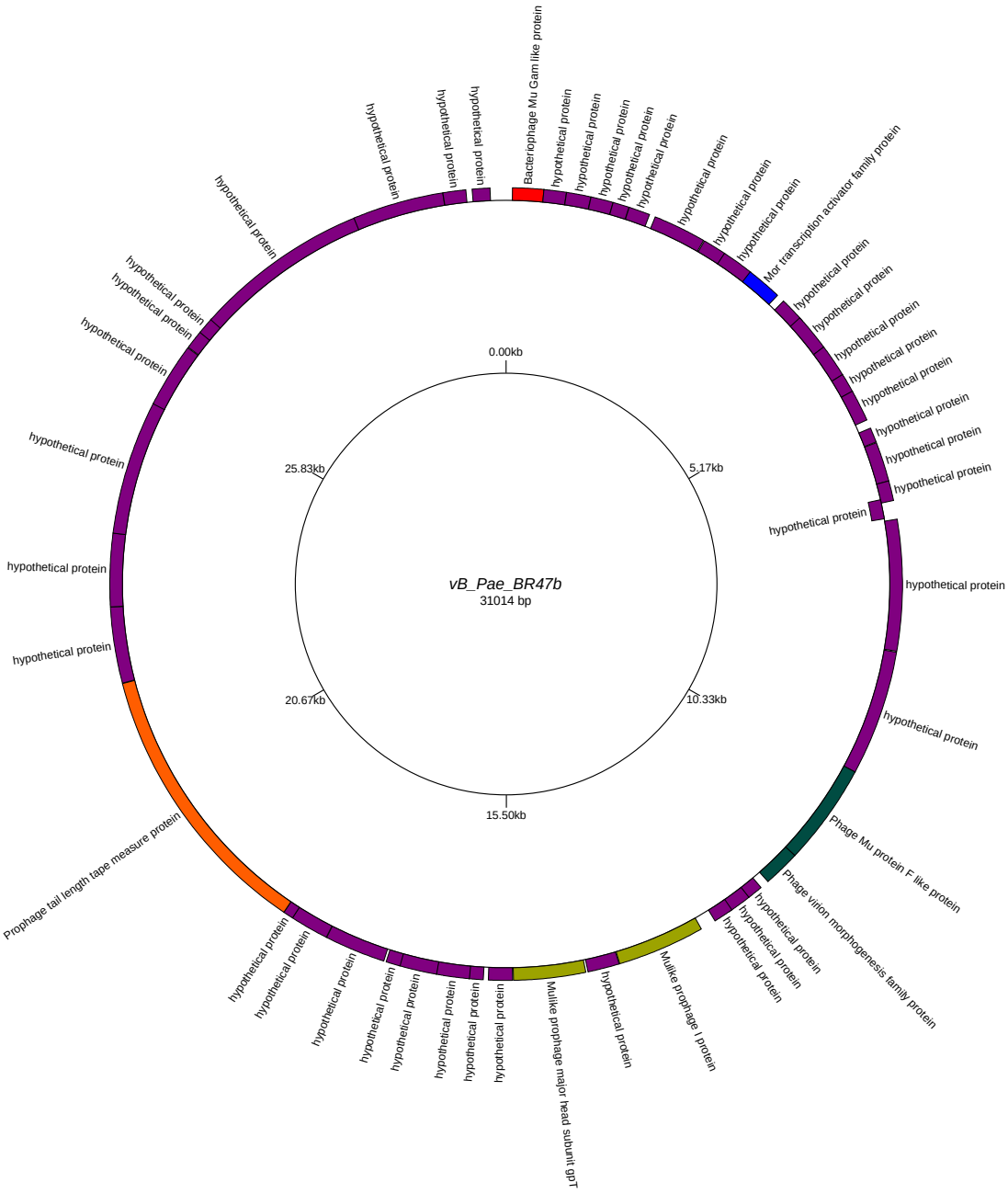


Figure 64: Genome map of vB_Pae_BR47b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR50a genome size was 38.3 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 65. The total number of genes identified was 58 of which 26 were identified with putative functions these can be seen in Table 65. Blastn on viruses only database showed a 1.9 kb / 1.9 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR50a had a 61.3 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
688	359	Reverse	HNH endonuclease
1395	883	Reverse	Bacteriophage lysis protein
2012	1395	Reverse	Chitinase class I
2341	2009	Reverse	Phage holin family (Lysis protein S)
2956	2426	Reverse	Phage regulatory protein Rha (Phage_pRha)
4120	3731	Reverse	Phage antitermination protein Q
5789	4392	Reverse	Replicative DNA helicase
6601	5786	Reverse	DNA replication protein DnaC
10862	10557	Reverse	Helix-turn-helix
10972	11631	Forward	LexA repressor
11780	12166	Forward	Bacterial regulatory proteins, luxR family
12478	13263	Forward	BRO family, N-terminal domain
13338	13568	Forward	Arc-like DNA binding domain protein
18046	19029	Forward	Transposase DDE domain protein
23299	19634	Reverse	Carbohydrate binding domain protein
23928	23356	Reverse	Bacteriophage lambda tail assembly protein I
25437	24679	Reverse	NlpC/P60 family protein
26186	25440	Reverse	Phage minor tail protein L
26521	26183	Reverse	Phage minor tail protein
29796	26521	Reverse	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
32494	32168	Reverse	Phage head-tail joining protein

32817	32494	Reverse	Phage gp6-like head-tail connector protein
34286	33072	Reverse	Phage capsid family protein
34927	34283	Reverse	Caudovirus prohead protease
36134	34911	Reverse	Phage portal protein
37816	36137	Reverse	Phage Terminase

Table 65: Functional genes identified for vB_Pae_BR50a.

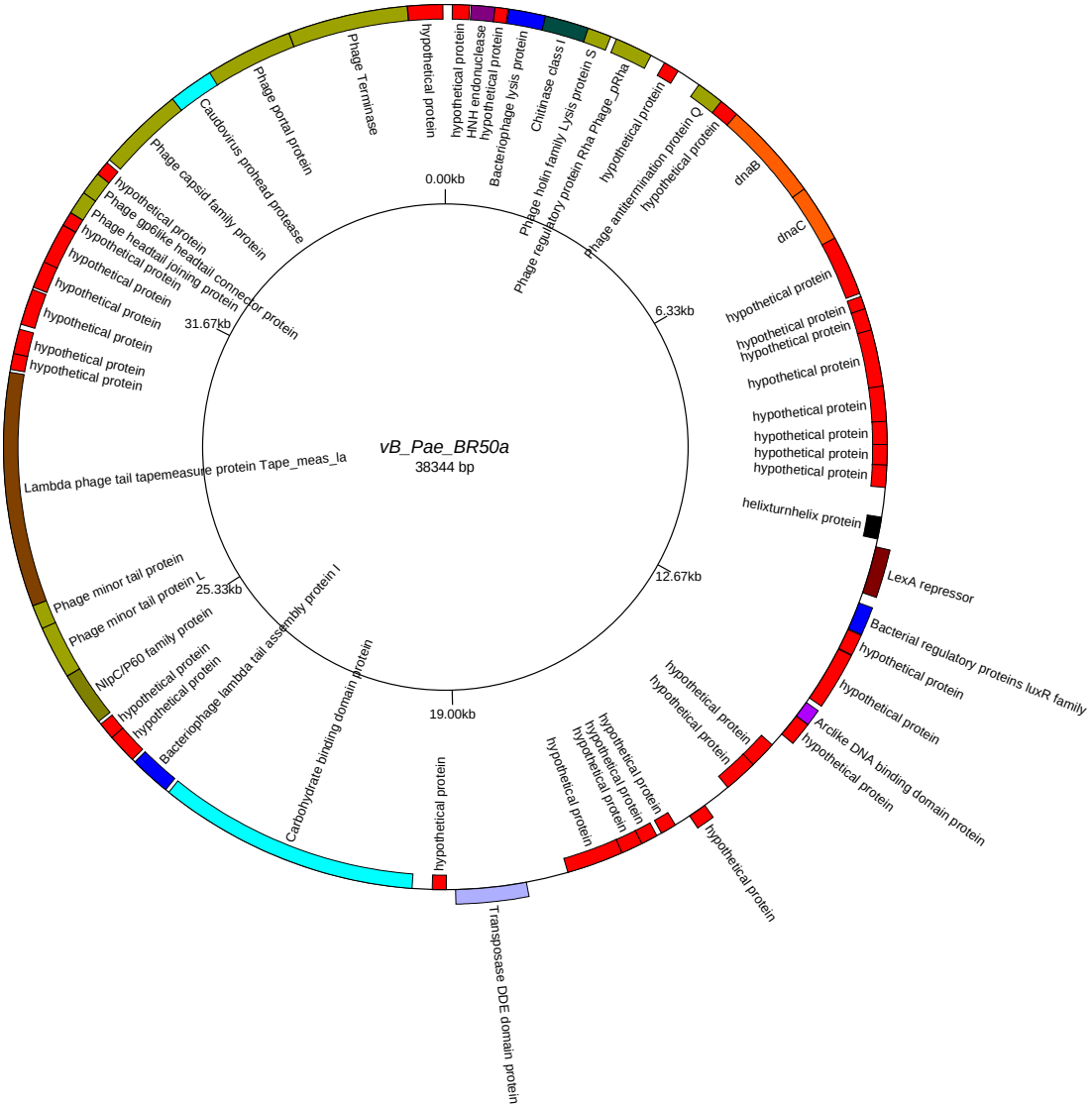


Figure 65: Genome map of vB_Pae_BR50a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR50b genome size was 48.2 kb, the phage was assembled using SPAdes and has a mean coverage of 123. The genome map of this phage is shown in Figure 66. The total number of genes identified was 81 of which 16 were identified with putative functions these can be seen in Table 66. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_BR50b had a 62.8 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
2620	2285	Reverse	LytTr DNA-binding domain protein
3243	2617	Reverse	YqaJ-like viral recombinase domain protein
3999	3247	Reverse	ERF superfamily protein
9604	8939	Reverse	putative HTH-type transcriptional regulator
9894	10397	Forward	Bacteriophage CII protein
10401	10805	Forward	T5orf172 domain protein
23219	19659	Reverse	Prophage tail length tape measure protein
27159	26251	Reverse	Mu-like prophage major head subunit gpT
28701	27595	Reverse	Mu-like prophage I protein
30187	29720	Reverse	Phage virion morphogenesis family protein
31473	30187	Reverse	Phage Mu protein F like protein
38391	37942	Reverse	Mor transcription activator family protein
41713	41195	Reverse	Bacteriophage Mu Gam like protein
45920	43926	Reverse	Mu DNA-binding domain protein
46978	46622	Reverse	DNA-binding transcriptional regulator Nlp
47148	47861	Forward	HTH-type transcriptional regulator PrtR

Table 66: Functional genes identified for vB_Pae_BR50b.

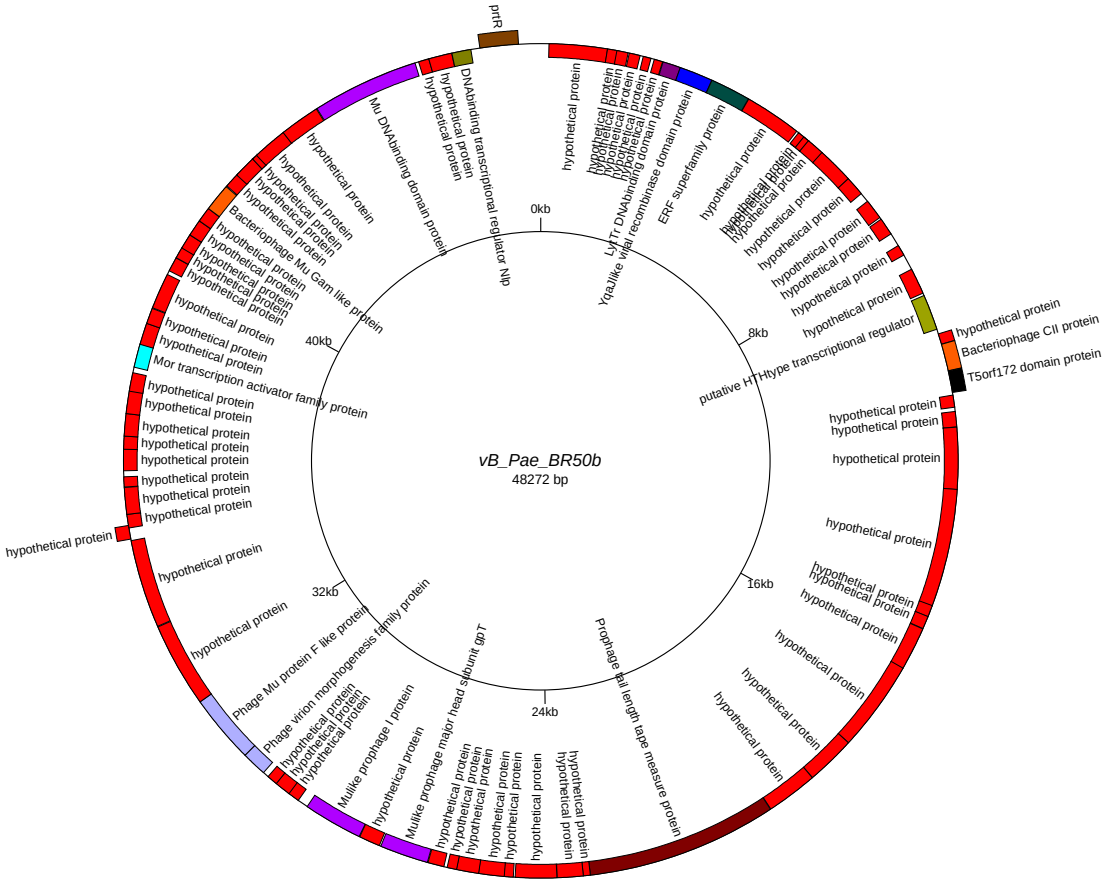


Figure 66: Genome map of vB_Pae_BR50b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR51a genome size was 38.3 kb, the phage was assembled using SPAdes and had a mean coverage of 46. The genome map of this phage is shown in Figure 67. The total number of genes identified was 58 of which 26 were identified with putative functions these can be seen in Table 67. Blastn on viruses only database showed a 1.9 kb / 1.9 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR51a had a 61.4 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
550	2229	Forward	Phage Terminase
2232	3455	Forward	Phage portal protein
3439	4083	Forward	Caudovirus prohead protease
4080	5294	Forward	Phage capsid family protein
5549	5872	Forward	Phage gp6-like head-tail connector protein
5872	6198	Forward	Phage head-tail joining protein
8570	11845	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
11845	12183	Forward	Phage minor tail protein
12180	12926	Forward	Phage minor tail protein L
12929	13687	Forward	NlpC/P60 family protein
14438	15010	Forward	Bacteriophage lambda tail assembly protein I
15067	18732	Forward	Carbohydrate binding domain protein
20320	19337	Reverse	Transposase DDE domain protein
25028	24798	Reverse	Arc-like DNA binding domain protein
25888	25103	Reverse	BRO family, N-terminal domain
26586	26200	Reverse	Bacterial regulatory proteins, luxR family
27394	26735	Reverse	LexA repressor
27504	27809	Forward	Helix-turn-helix
31765	32580	Forward	DNA replication protein DnaC
32577	33974	Forward	Replicative DNA helicase
34246	34635	Forward	Phage antitermination protein Q

35410	35940	Forward	Phage regulatory protein Rha (Phage_pRha)
36025	36357	Forward	Phage holin family (Lysis protein S)
36354	36971	Forward	Chitinase class I
36971	37483	Forward	Bacteriophage lysis protein
37678	38007	Forward	HNH endonuclease

Table 67: Functional genes identified for vB_Pae_BR51a.

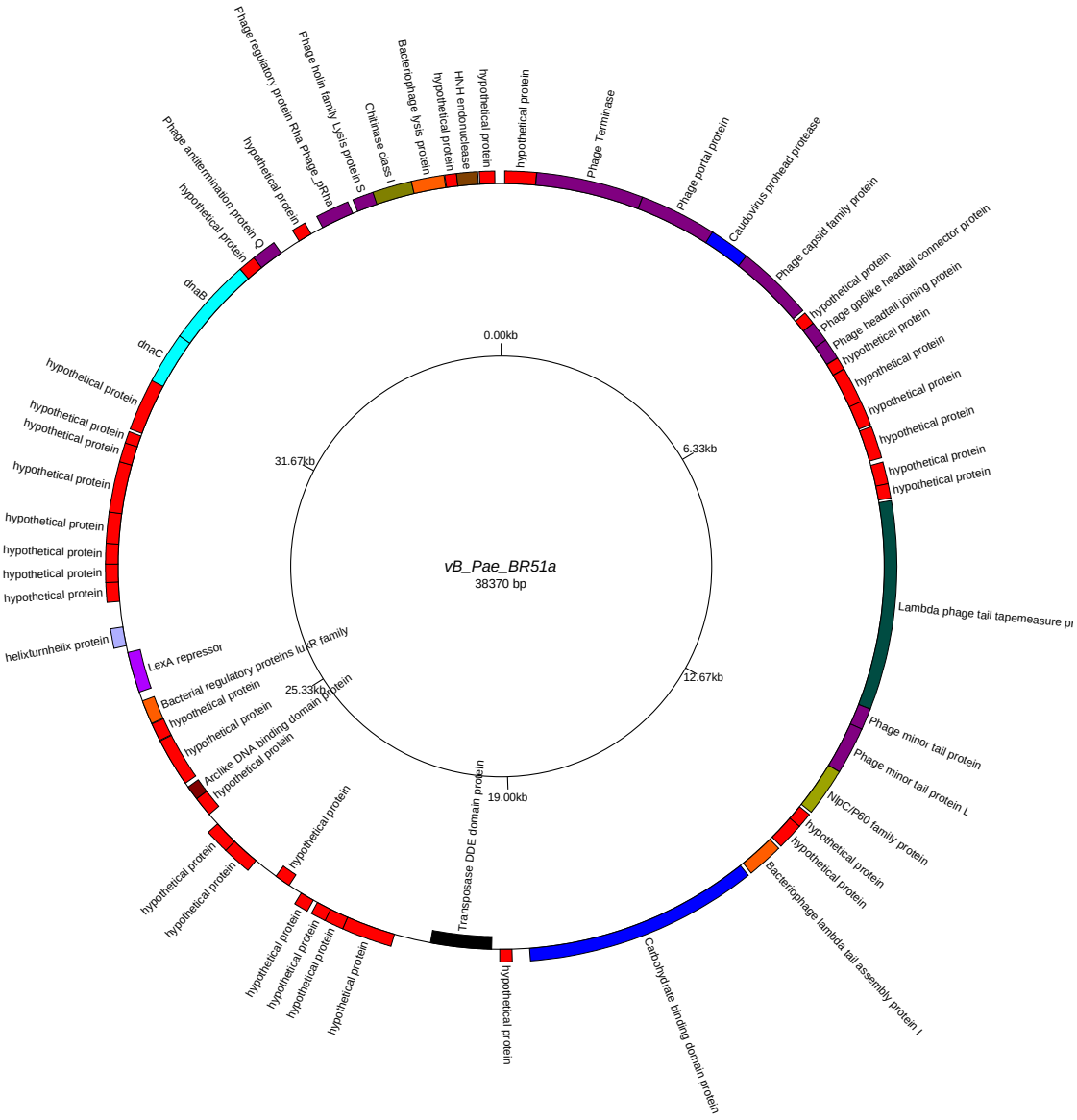


Figure 67: Genome map of vB_Pae_BR51a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR53a genome size was 50 kb, the phage was assembled using SPAdes and has a mean coverage of 506. The genome map of this phage is shown in Figure 67. The total number of genes identified was 72 of which 15 were identified with putative functions these can be seen in Table 67. Blastn on viruses only database showed an 8.6 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_BR53a had a 61.8 % G-C content. No tRNA gene was identified however; CII and Ig-like domains were identified along with recombinase genes.

Start	End	Direction	Putative functional protein
2857	3171	Forward	Helix-turn-helix domain protein
3210	4166	Forward	Tyrosine recombinase XerC
5981	5352	Reverse	Chitinase class I
18224	17229	Reverse	Bacterial Ig-like domain (group 2)
24052	23129	Reverse	Phage Mu protein F like protein
26838	25585	Reverse	Phage terminase large subunit
27364	26822	Reverse	Terminase small subunit
29312	28869	Reverse	Endodeoxyribonuclease RusA
31159	30848	Reverse	Bacteriophage CII protein
31449	32114	Reverse	putative HTH-type transcriptional regulator
34835	35206	Forward	Carbon storage regulator homolog
36950	37858	Forward	YqaJ-like viral recombinase domain protein
37871	38761	Forward	recombination and repair protein RecT
39749	41491	Forward	chromosome segregation protein
42101	43891	Forward	C-5 cytosine-specific DNA methylase

Table 68: Functional genes identified for vB_Pae_BR53a.

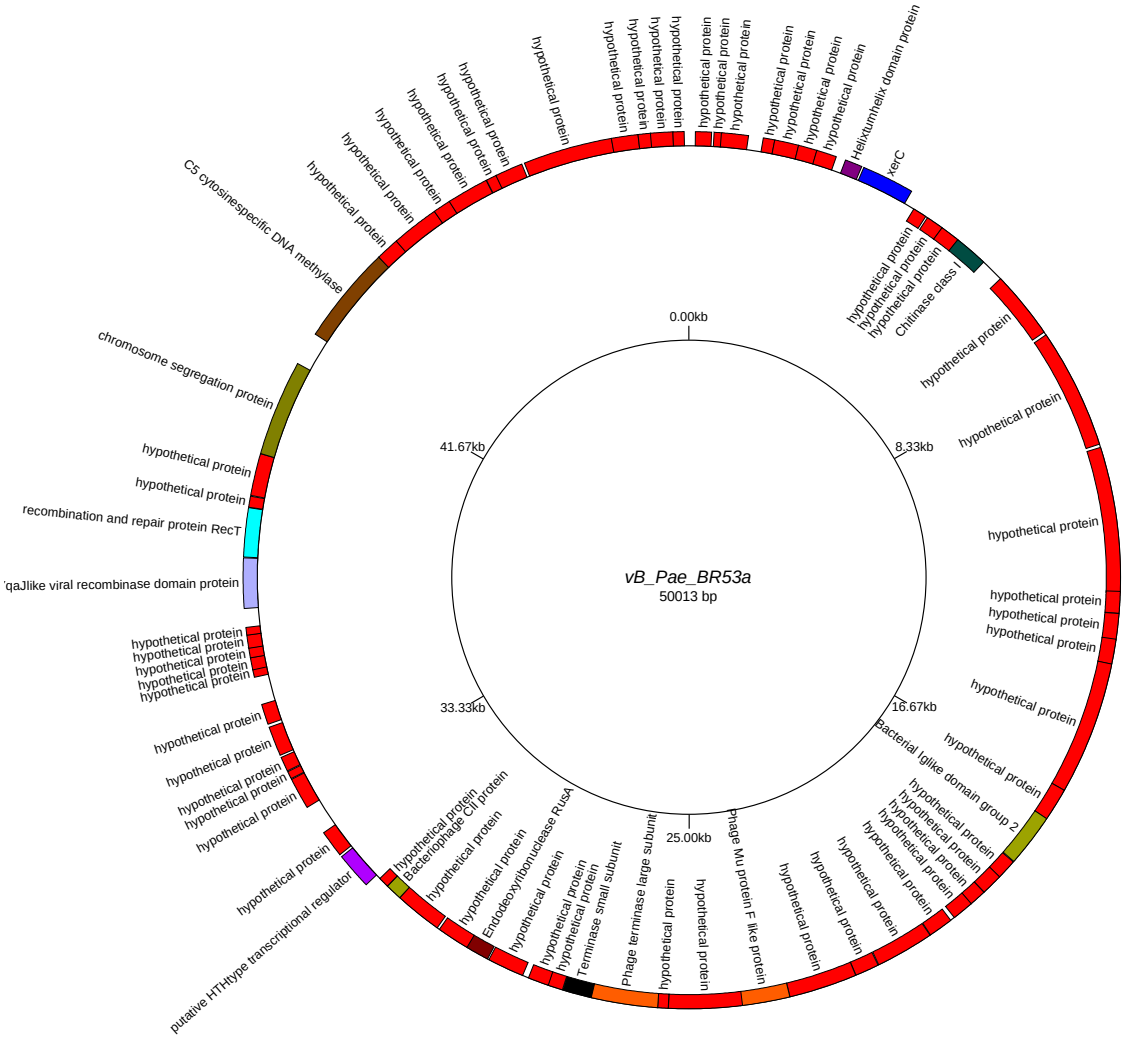


Figure 68: Genome map of vB_Pae_BR53a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR54a genome size was 42.4 kb, the phage was assembled using SPAdes and has a mean coverage of 226. The genome map of this phage is shown in Figure 69. The total number of genes identified was 65 of which 11 were identified with putative functions these can be seen in Table 69. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_BR54a had a 59.2 % G-C content. A tRNA gene was identified; CII and Ig-like domains were identified along with recombinase genes.

Start	End	Direction	Putative functional protein
6284	5274	Reverse	site-specific tyrosine recombinase XerC
6445	6358	Reverse	tRNA-Ser(cga)
7949	7515	Reverse	Lysozyme RrrD
8552	9451	Forward	BRO family, N-terminal domain
11592	9946	Reverse	D-glucuronyl C5-epimerase C-terminus
19047	18487	Reverse	AP2 domain protein
24128	23133	Reverse	Bacterial Ig-like domain (group 2)
30743	29013	Reverse	Phage Mu protein F like protein
33394	32276	Reverse	Phage terminase large subunit
38219	37572	Reverse	Bacteriophage Lambda NinG protein
42018	40633	Reverse	Replicative DNA helicase

Table 69: Functional genes identified for vB_Pae_BR54a.

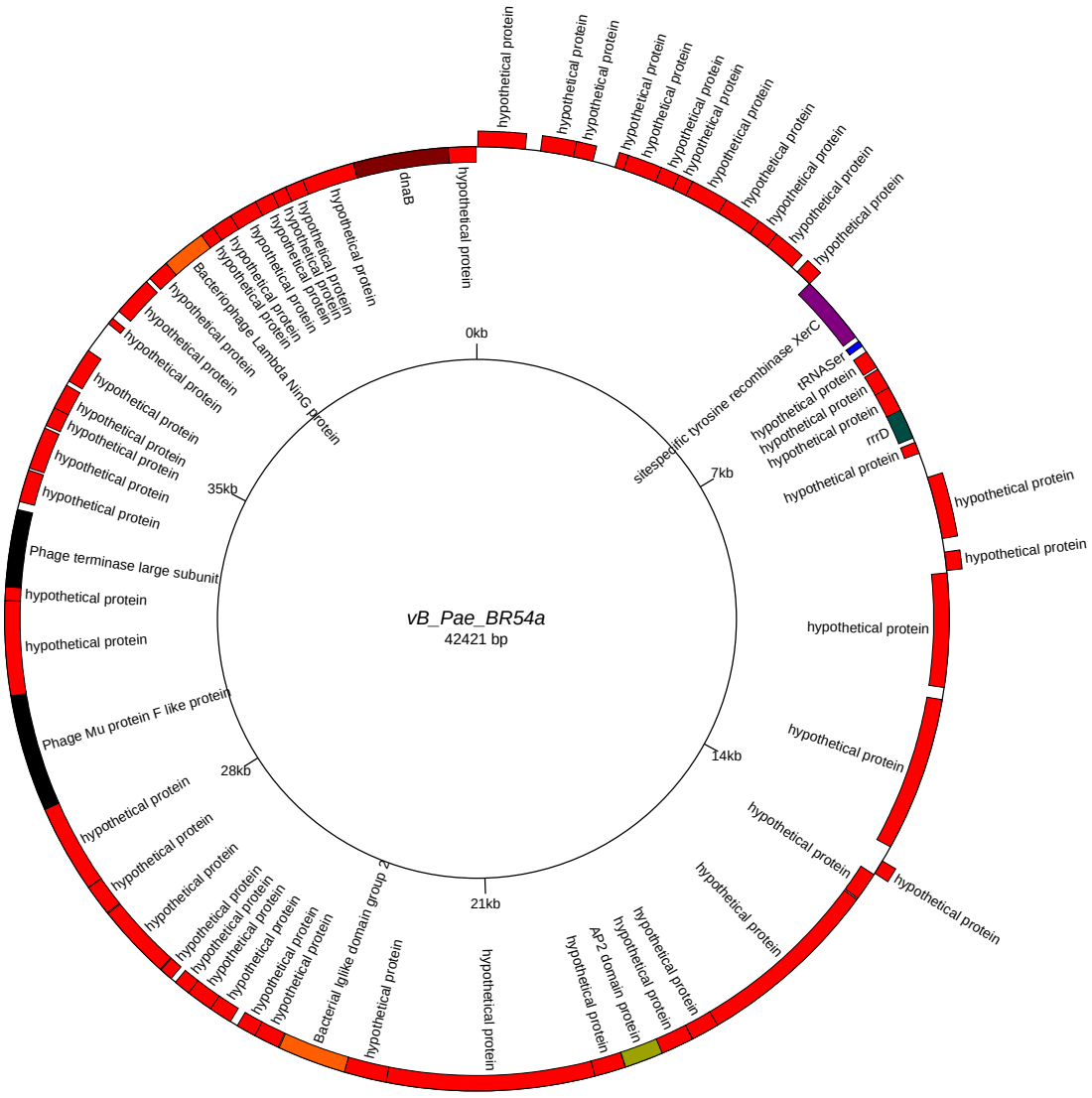


Figure 69: Genome map of vB_Pae_BR54a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR54b genome size was 39.6 kb, the phage was assembled using SPAdes and has a mean coverage of 227. The genome map of this phage is shown in Figure 70. The total number of genes identified was 65 of which 11 were identified with putative functions these can be seen in Table 70. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_BR54b had a 64.1 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
12692	9132	Reverse	Prophage tail length tape measure protein
16632	15724	Reverse	Mu-like prophage major head subunit gpT
18174	17068	Reverse	Mu-like prophage I protein
19660	19193	Reverse	Phage virion morphogenesis family protein
20946	19660	Reverse	Phage Mu protein F like protein
27864	27415	Reverse	Mor transcription activator family protein
31186	30668	Reverse	Bacteriophage Mu Gam like protein
35393	33399	Reverse	Mu DNA-binding domain protein
36451	36095	Reverse	DNA-binding transcriptional regulator Nlp
36621	37334	Forward	HTH-type transcriptional regulator PrtR
37784	38119	Forward	LytTr DNA-binding domain protein

Table 70: Functional genes identified for vB_Pae_BR54b.

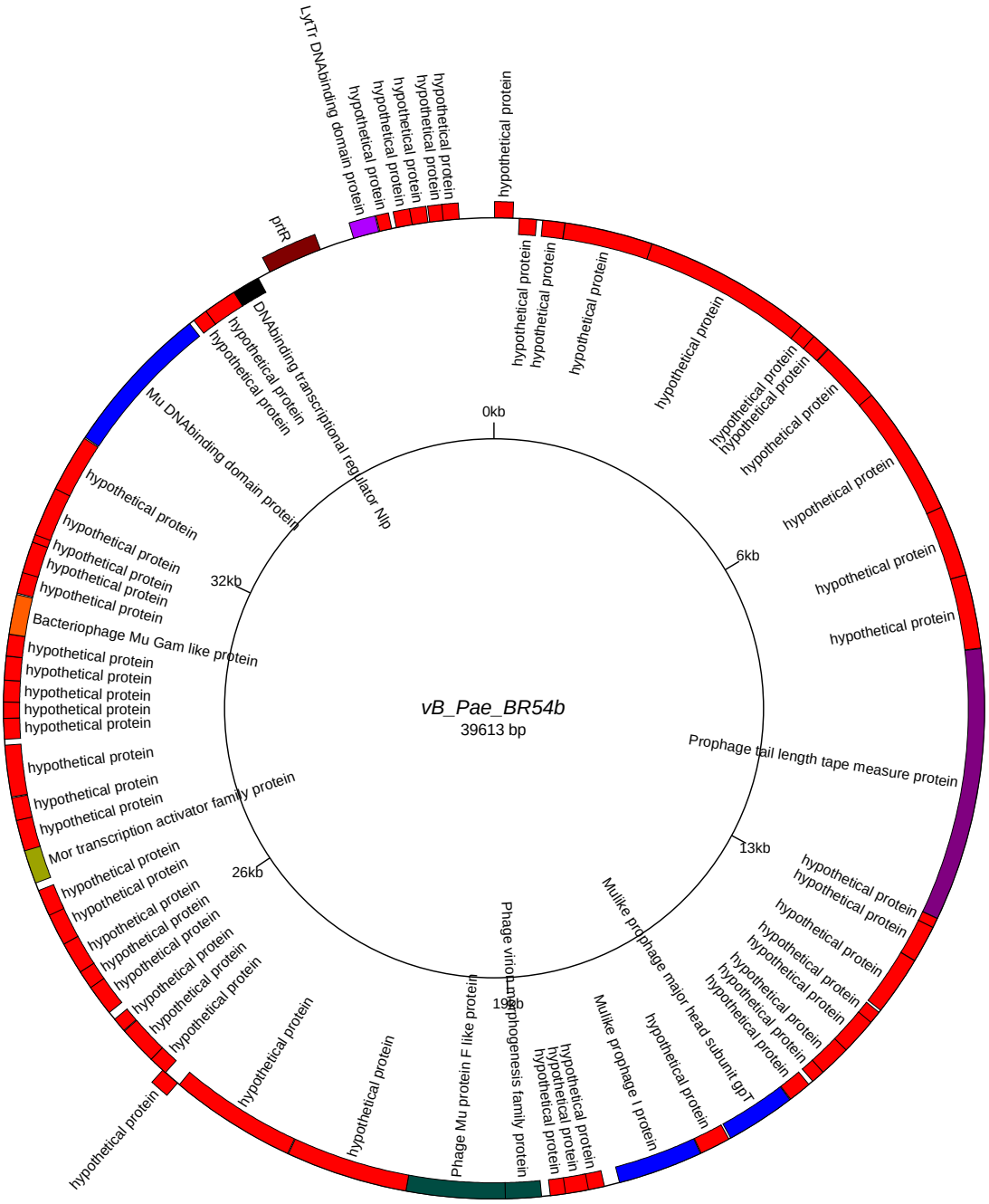


Figure 70: Genome map of vB_Pae_BR54b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR54c genome size was 38.2 kb, the phage was assembled using SPAdes and had a mean coverage of 35. The genome map of this phage is shown in Figure 71. The total number of genes identified was 58 of which 26 were identified with putative functions these can be seen in Table 71. Blastn on viruses only database showed a 1.9 kb / 1.9 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR54c had a 61.4 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
471	2150	Forward	Phage Terminase
2153	3376	Forward	Phage portal protein
3360	4004	Forward	Caudovirus prohead protease
4001	5215	Forward	Phage capsid family protein
5470	5793	Forward	Phage gp6-like head-tail connector protein
5793	6119	Forward	Phage head-tail joining protein
8491	11766	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
11766	12104	Forward	Phage minor tail protein
12101	12847	Forward	Phage minor tail protein L
12850	13608	Forward	NlpC/P60 family protein
14359	14931	Forward	Bacteriophage lambda tail assembly protein I
14988	18653	Forward	Carbohydrate binding domain protein
20241	19258	Reverse	Transposase DDE domain protein
24949	24719	Reverse	Arc-like DNA binding domain protein
25809	25024	Reverse	BRO family, N-terminal domain
26507	26121	Reverse	Bacterial regulatory proteins, luxR family
27315	26656	Reverse	LexA repressor
27425	27730	Forward	Helix-turn-helix
31686	32501	Forward	DNA replication protein DnaC
32498	33895	Forward	Replicative DNA helicase
34167	34556	Forward	Phage antitermination protein Q

35331	35861	Forward	Phage regulatory protein Rha (Phage_pRha)
35946	36278	Forward	Phage holin family (Lysis protein S)
36275	36892	Forward	Chitinase class I
36892	37404	Forward	Bacteriophage lysis protein
37599	37928	Forward	HNH endonuclease

Table 71: Functional genes identified for vB_Pae_BR54c.

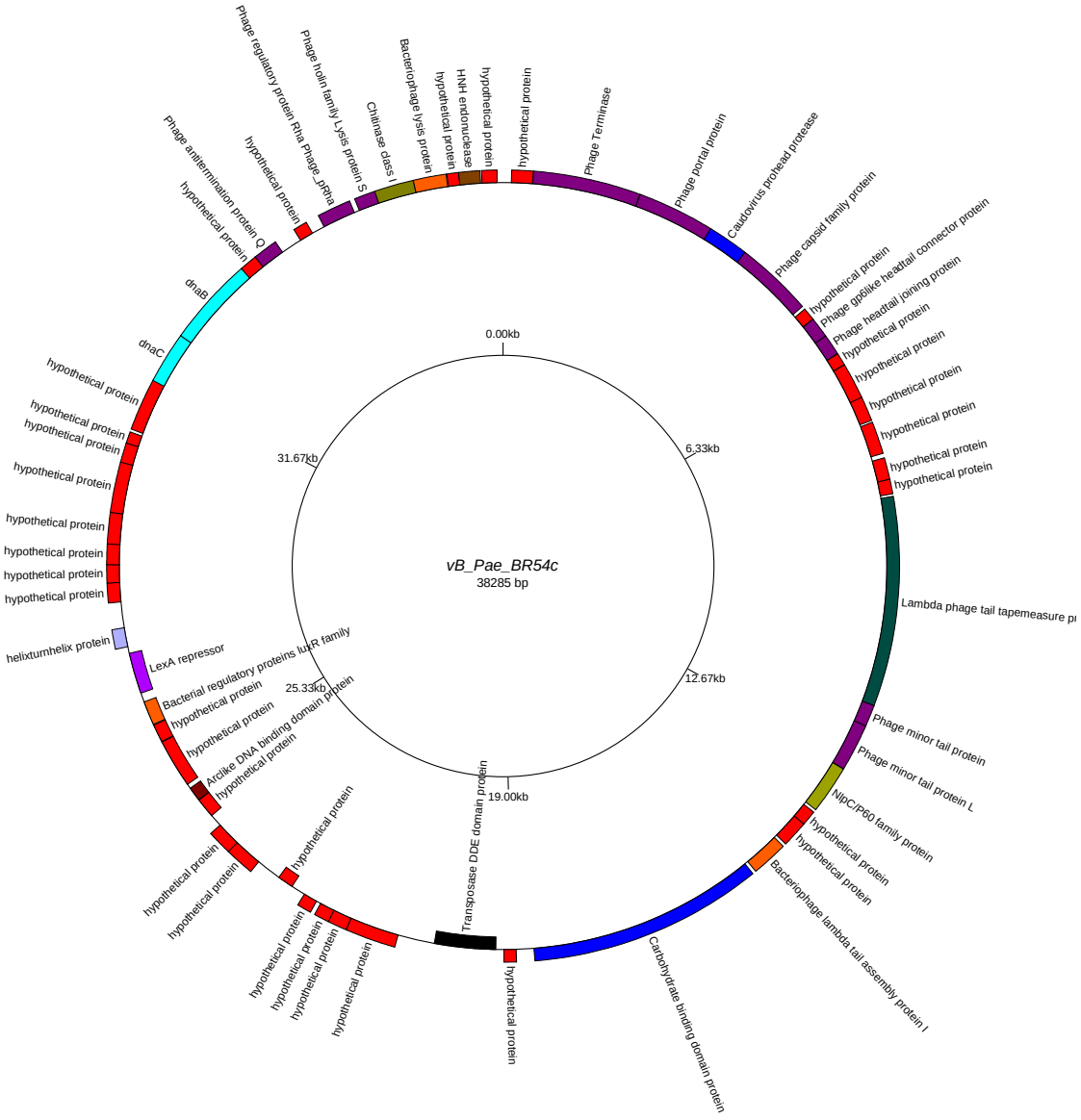


Figure 71: Genome map of vB_Pae_BR54c produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR55a genome size was 37.5 kb, the phage was assembled using SPAdes and has a mean coverage of 88. The genome map of this phage is shown in Figure 72. The total number of genes identified was 57 of which 9 were identified with putative functions these can be seen in Table 72. Blastn on viruses only database showed a 27.1 kb / 27.8 kb match with *Pseudomonas* phage JBD24 (GenBank: NC_020203.1). The phage vB_Pae_BR55a had a 64.2 % G-C content. No tRNA gene was identified however; several genes associated with structural proteins were identified.

Start	End	Direction	Putative functional protein
16669	15755	Reverse	Mu-like prophage major head subunit gpT
17770	16673	Reverse	Mu-like prophage I protein
19275	18808	Reverse	Phage virion morphogenesis family protein
20561	19275	Reverse	Phage Mu protein F like protein
27468	27019	Reverse	Mor transcription activator family protein
34924	32855	Reverse	Mu DNA-binding domain protein
36443	36087	Reverse	DNA-binding transcriptional regulator Nlp
36684	37343	Forward	putative HTH-type transcriptional regulator
37815	37450	Reverse	4-hydroxybenzoate transporter PcaK

Table 72: Functional genes identified for vB_Pae_BR55a.

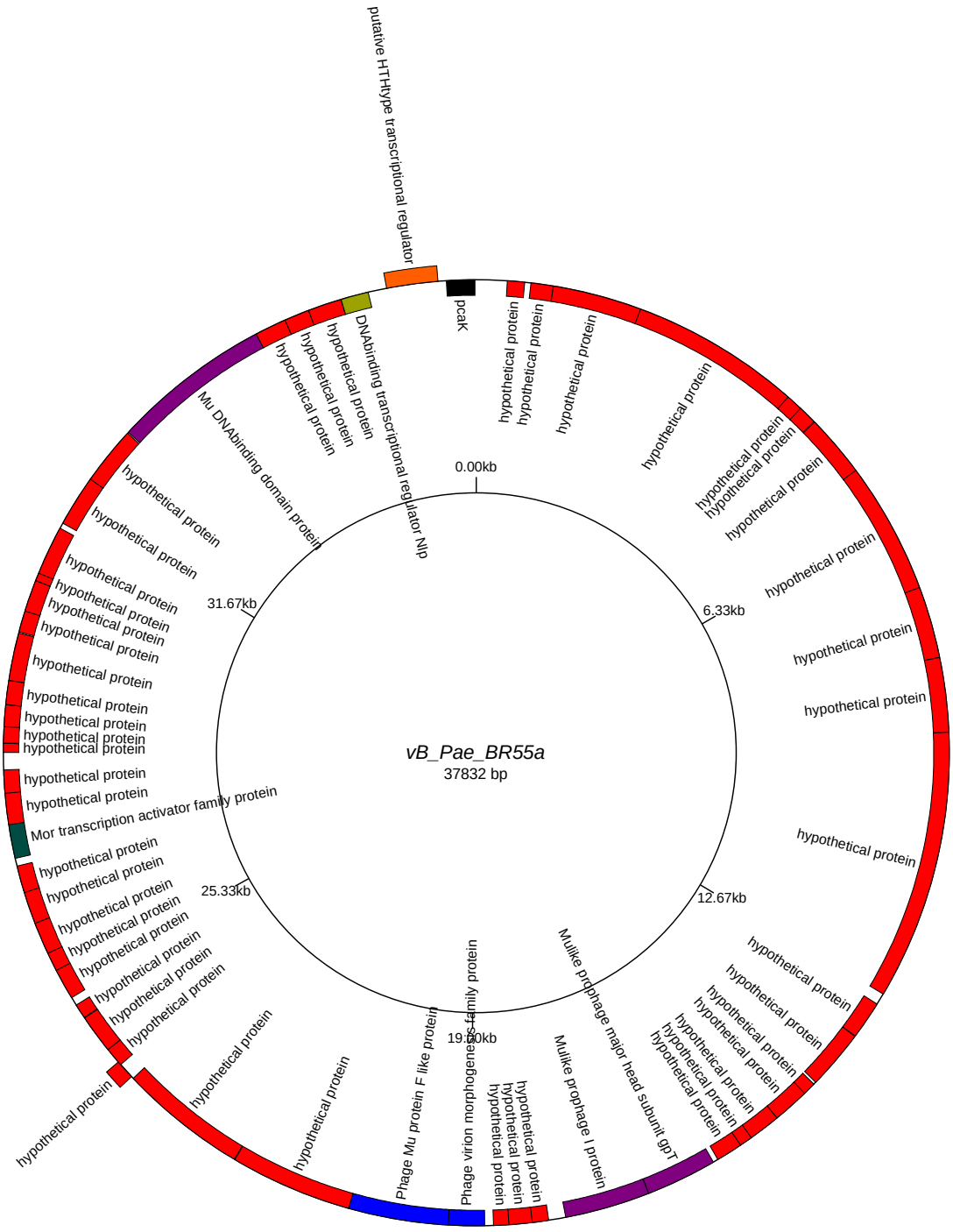
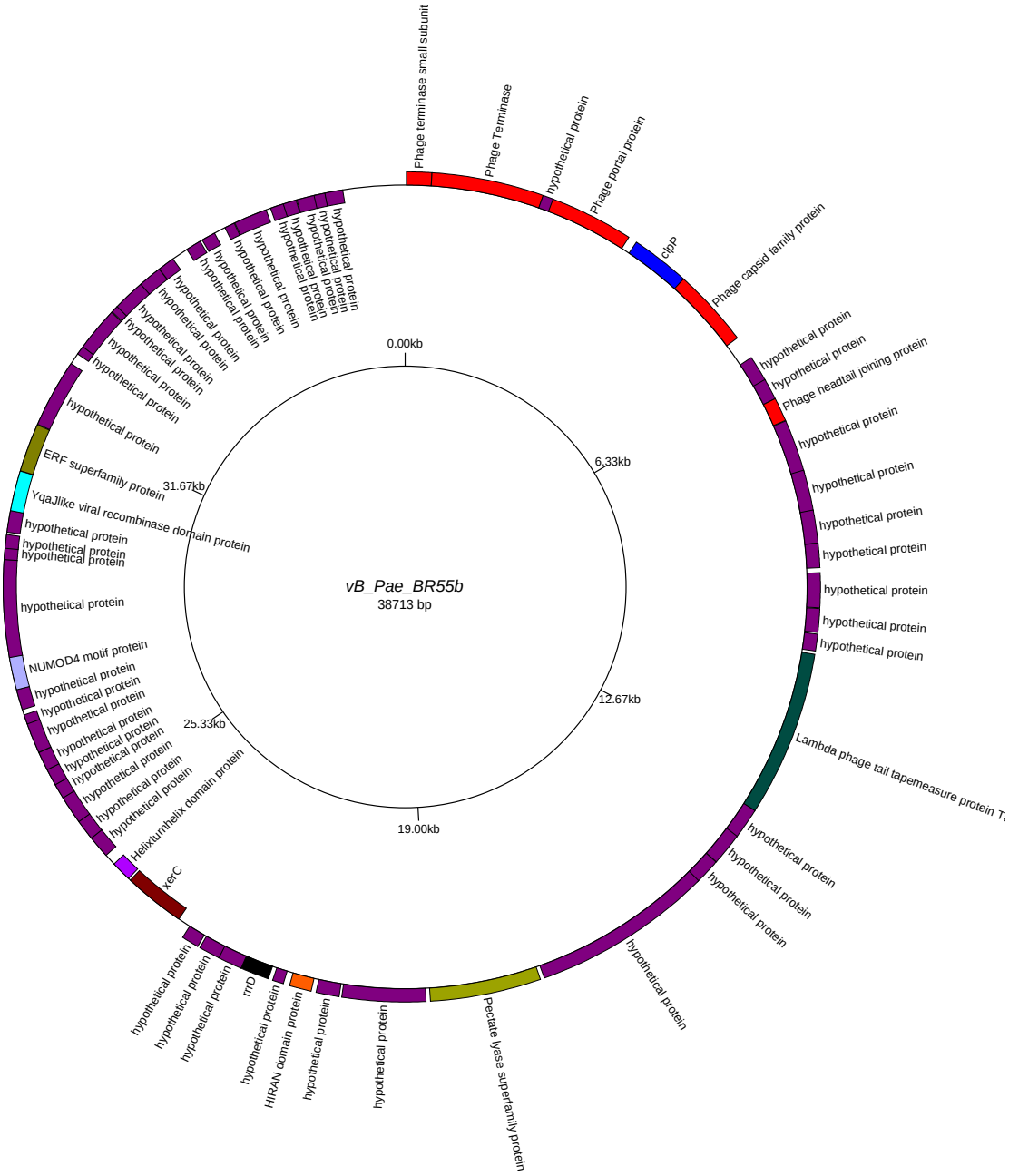


Figure 72: Genome map of vB_Pae_BR55a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR55b genome size was 38.7 kb, the phage was assembled using SPAdes and has a mean coverage of 4.7. The genome map of this phage is shown in Figure 73. The total number of genes identified was 64 of which 15 were identified with putative functions these can be seen in Table 73. Blastn on viruses only database showed a 6.3 kb / 7 kb match with *Pseudomonas* phage vB_PaeS_PMG1 (GenBank: NC_016765.1). The phage vB_Pae_BR55b had a 58.4 % G-C content. No tRNA gene was identified however; genes associated with structural, lysogenic and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
17	397	Forward	Phage terminase, small subunit
399	2090	Forward	Phage Terminase
2244	3506	Forward	Phage portal protein
3638	4528	Forward	ATP-dependent Clp protease proteolytic subunit
4525	5712	Forward	Phage capsid family protein
6777	7133	Forward	Phage head-tail joining protein
10683	13187	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
17312	18982	Forward	Pectate lyase superfamily protein
20765	21097	Forward	HIRAN domain protein
21428	21862	Forward	Lysozyme RrrD
23995	23039	Reverse	Tyrosine recombinase XerC
24348	24034	Reverse	Helix-turn-helix domain protein
27954	27439	Reverse	NUMOD4 motif protein
30838	30212	Reverse	YqaJ-like viral recombinase domain protein
31594	30842	Reverse	ERF superfamily protein

Table 73: Functional genes identified for vB_Pae_BR55b.


vB_Pae_BR55b
38713 bp
0.00kb
6.33kb
12.67kb
19.00kb
25.33kb
31.67kb
Phage terminase small subunit
Phage Terminase
hypothetical protein
Phage portal protein
clpP
Phage capsid family protein
hypothetical protein
hypothetical protein
Phage headtail joining protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Lambda phage tail tapemeasure protein T
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Pectate lyase superfamily protein
hypothetical protein
hypothetical protein
HIRAN domain protein
hypothetical protein
rnlD
hypothetical protein
hypothetical protein
hypothetical protein
XerC
Helixturnhelix domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
NUMOD4 motif protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
YqaJlike viral recombinase domain protein
ERF superfamily protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Pae_BR56a genome size was 38.4 kb, the phage was assembled using SPAdes and had a mean coverage of 26. The genome map of this phage is shown in Figure 74. The total number of genes identified was 58 of which 26 were identified with putative functions these can be seen in Table 74. Blastn on viruses only database showed a 1.9 kb / 1.9 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR56a had a 61.3 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
688	359	Reverse	HNH endonuclease
1395	883	Reverse	Bacteriophage lysis protein
2012	1395	Reverse	Chitinase class I
2341	2009	Reverse	Phage holin family (Lysis protein S)
2956	2426	Reverse	Phage regulatory protein Rha (Phage_pRha)
4120	3731	Reverse	Phage antitermination protein Q
5789	4392	Reverse	Replicative DNA helicase
6601	5786	Reverse	DNA replication protein DnaC
10862	10557	Reverse	Helix-turn-helix
10972	11631	Forward	LexA repressor
11780	12166	Forward	Bacterial regulatory proteins, luxR family
12478	13263	Forward	BRO family, N-terminal domain
13338	13568	Forward	Arc-like DNA binding domain protein
18046	19029	Forward	Transposase DDE domain protein
23299	19634	Reverse	Carbohydrate binding domain protein
23928	23356	Reverse	Bacteriophage lambda tail assembly protein I
25437	24679	Reverse	NlpC/P60 family protein
26186	25440	Reverse	Phage minor tail protein L
26521	26183	Reverse	Phage minor tail protein
29796	26521	Reverse	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
32494	32168	Reverse	Phage head-tail joining protein

32817	32494	Reverse	Phage gp6-like head-tail connector protein
34286	33072	Reverse	Phage capsid family protein
34927	34283	Reverse	Caudovirus prohead protease
36134	34911	Reverse	Phage portal protein
37816	36137	Reverse	Phage Terminase

Table 74: Functional genes identified for vB_Pae_BR56a.

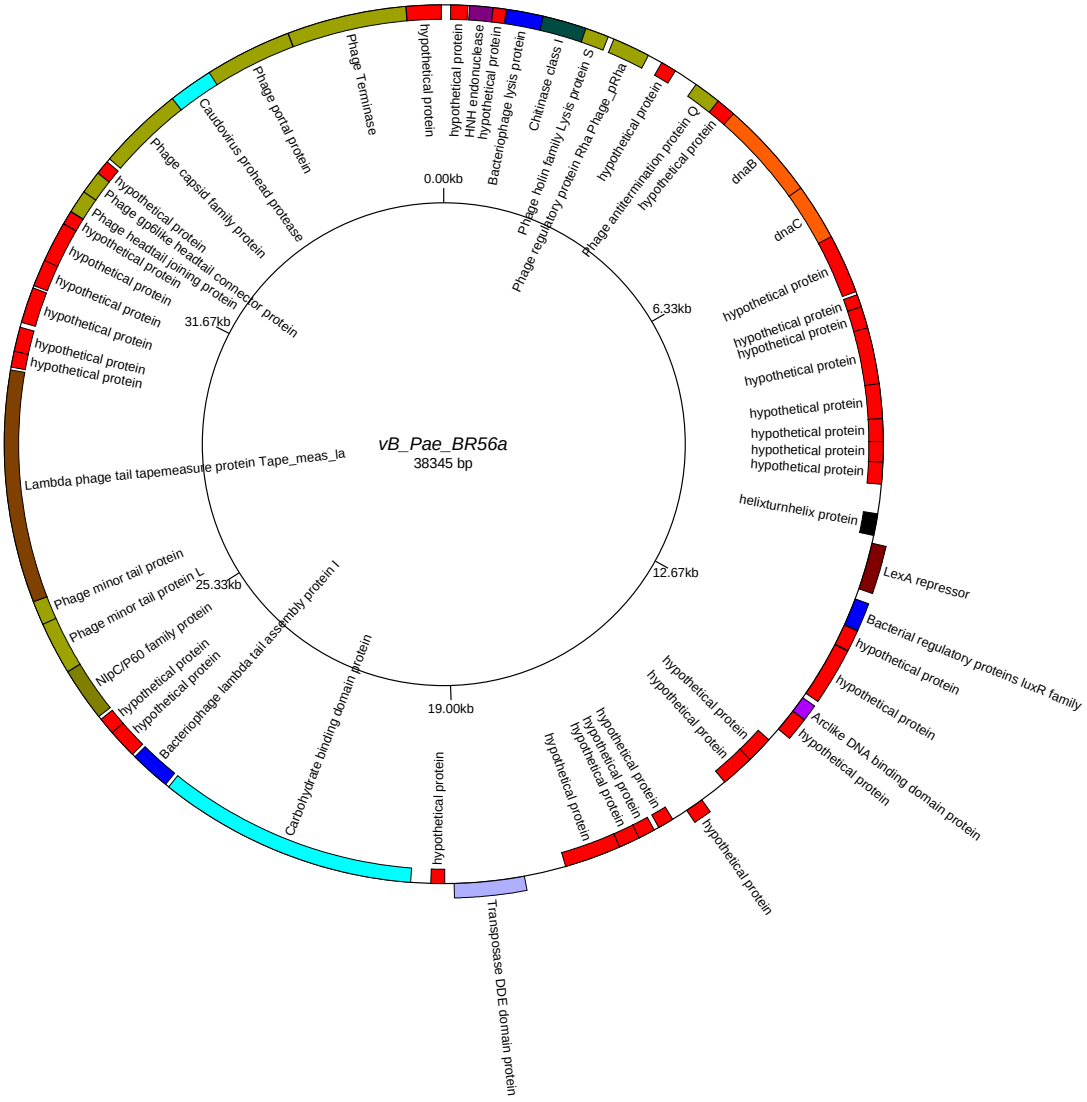


Figure 74: Genome map of vB_Pae_BR56a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR59a genome size was 61.7 kb, the phage was assembled using SPAdes and has a mean coverage of 457. The genome map of this phage is shown in Figure 74. The total number of genes identified was 62 of which 14 were identified with putative functions these can be seen in Table 74. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66 (GenBank: KC262634.1). The phage vB_Pae_BR59a had a 63.7 % G-C content. No tRNA gene was identified however; a CPS-53 integrase like gene was identified.

Start	End	Direction	Putative functional protein
4605	5102	Forward	Single-stranded DNA-binding protein
5129	6349	Forward	Recombination-associated protein RdgC
6346	6555	Forward	LexA repressor
6700	9330	Forward	DNA methylase
9327	11135	Forward	C-5 cytosine-specific DNA methylase
13664	13897	Forward	Response regulator inhibitor for tor operon
15133	13898	Reverse	Putative prophage CPS-53 integrase
18738	18199	Reverse	Phage lysozyme
47162	46290	Reverse	RyR domain protein
53961	52678	Reverse	Phage terminase large subunit
55759	55145	Reverse	Bacteriophage Lambda NinG protein
58365	58165	Reverse	Cro
58473	59270	Forward	HTH-type transcriptional regulator PrtR
60689	61060	Forward	Carbon storage regulator homolog

Table 75: Functional genes identified for vB_Pae_BR59a.

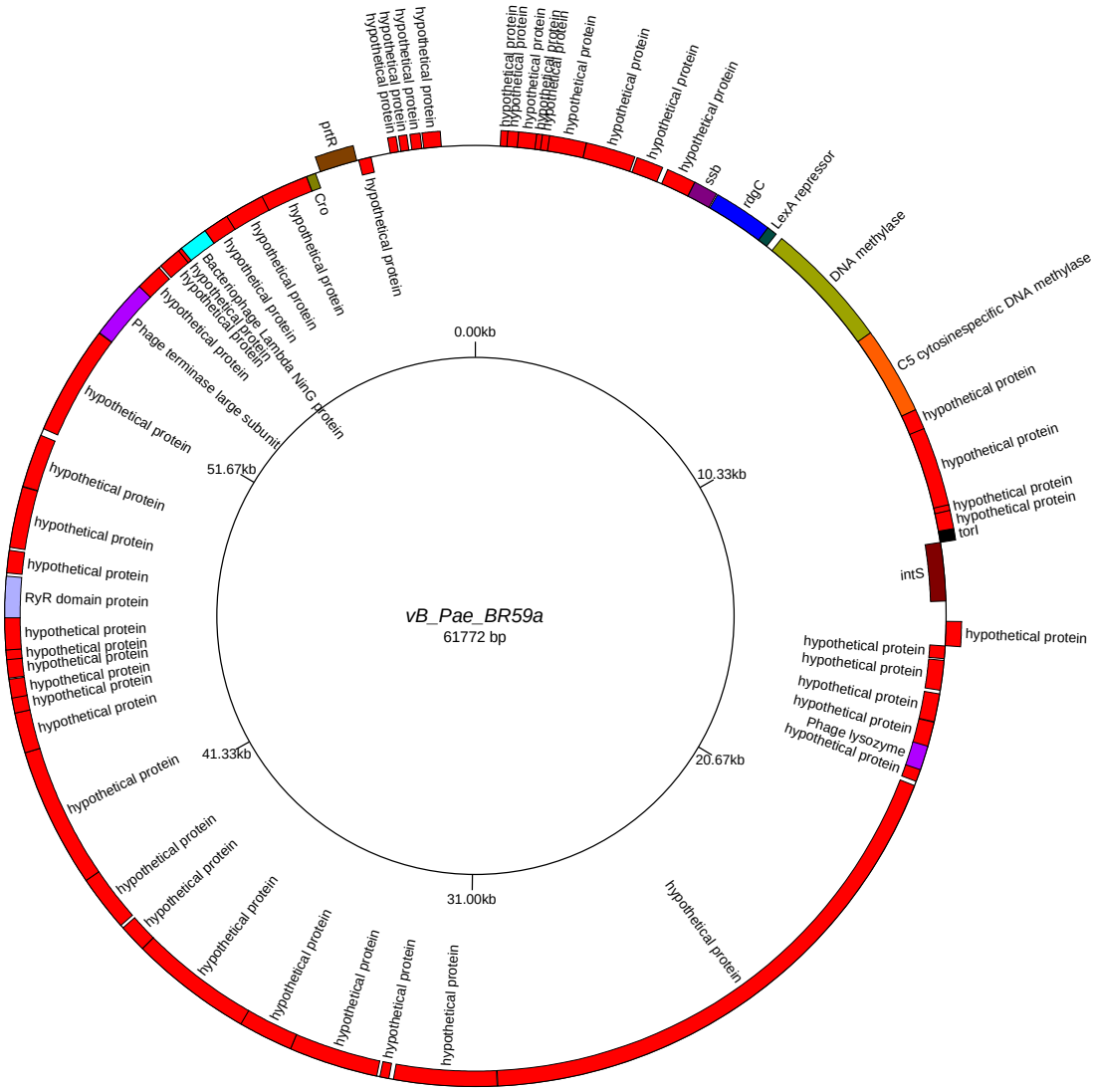


Figure 75: Genome map of vB_Pae_BR59a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR61a genome size was 38.1 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 76. The total number of genes identified was 60 of which 13 were identified with putative functions these can be seen in Table 76. Blastn on viruses only database showed an 8 kb / 8.6 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR61a had a 61.3 % G-C content. No tRNA gene was identified however; genes associated with structural, lysogenic and regulatory proteins were identified, including an integrase putative protein.

Start	End	Direction	Putative functional protein
7623	6571	Reverse	Phage integrase family protein
10911	10129	Reverse	Phage regulatory protein Rha (Phage_pRha)
12316	11930	Reverse	Bacterial regulatory proteins, luxR family
14190	13399	Reverse	HTH-type transcriptional regulator PrtR
21913	22302	Forward	Phage antitermination protein Q
22841	23368	Forward	Phage regulatory protein Rha (Phage_pRha)
23453	23785	Forward	Phage holin family (Lysis protein S)
23782	24399	Forward	Chitinase class I
24635	25105	Forward	Bacteriophage lysis protein
25962	26507	Forward	Phage DNA packaging protein Nu1
26479	28443	Forward	Phage terminase large subunit (GpA)
28649	30295	Forward	Phage portal protein, lambda family
30267	32348	Forward	ATP-dependent Clp protease proteolytic subunit

Table 76: Functional genes identified for vB_Pae_BR61a.

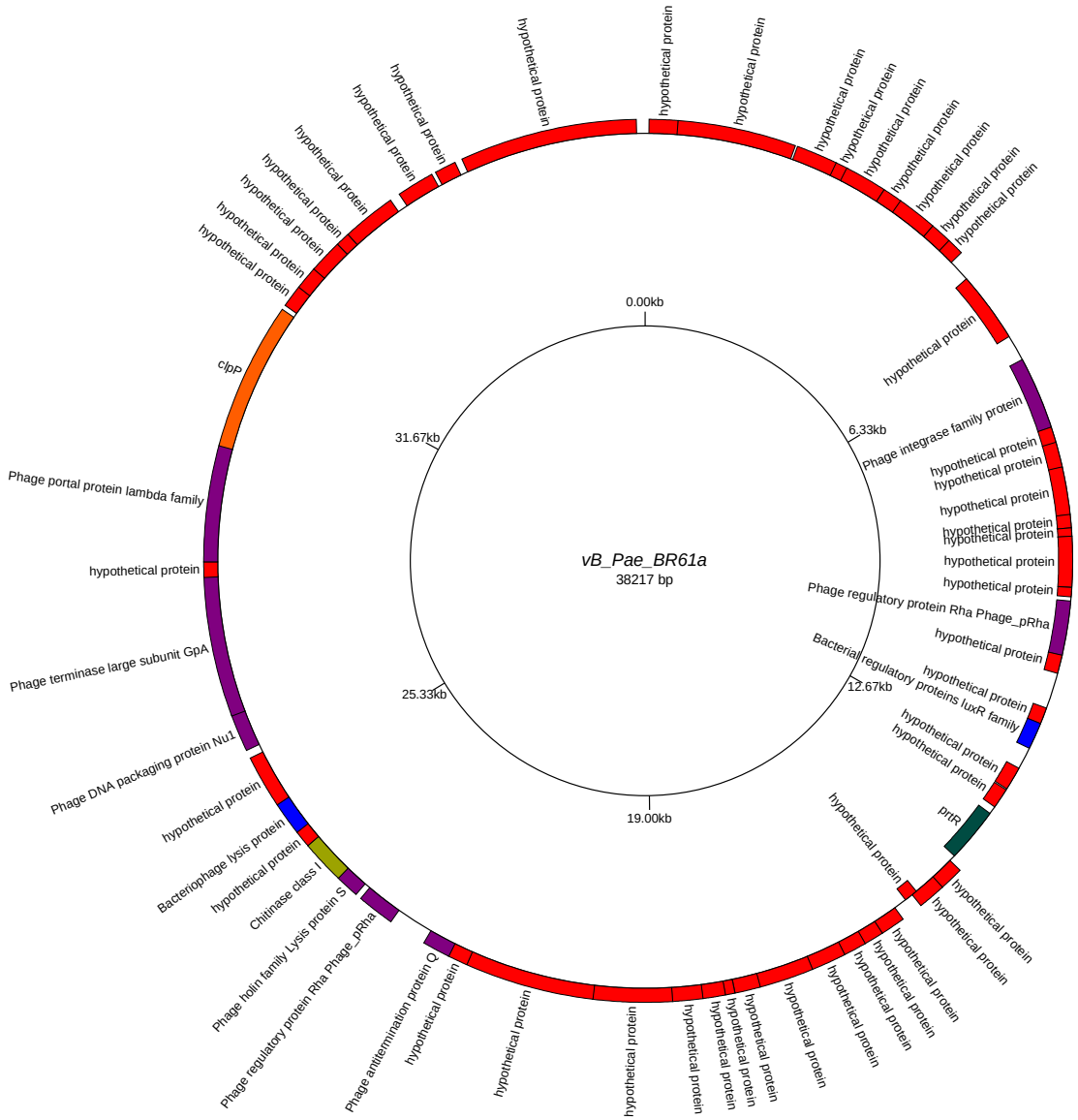


Figure 76: Genome map of vB_Pae_BR61a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR62a genome size was 40.4 kb, the phage was assembled using SPAdes and has a mean coverage of 227. The genome map of this phage is shown in Figure 77. The total number of genes identified was 61 of which 13 were identified with putative functions these can be seen in Table 77. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR62a had a 61.6 % G-C content. No tRNA gene was identified however; a *Cro* gene and phage integrase like gene was identified.

Start	End	Direction	Putative functional protein
1291	821	Reverse	Bacteriophage lysis protein
2144	1527	Reverse	Chitinase class I
2473	2141	Reverse	Phage holin family (Lysis protein S)
3085	2558	Reverse	Phage regulatory protein Rha (Phage_pRha)
4013	3624	Reverse	Phage antitermination protein Q
11629	11309	Reverse	Cro
11732	12520	Forward	HTH-type transcriptional regulator PrtR
14274	15041	Forward	Phage regulatory protein Rha (Phage_pRha)
17450	18502	Forward	Phage integrase family protein
36074	33993	Reverse	ATP-dependent Clp protease proteolytic subunit
37692	36046	Reverse	Phage portal protein, lambda family
39862	37898	Reverse	Phage terminase large subunit (GpA)
40418	39834	Reverse	Phage DNA packaging protein Nu1

Table 77: Functional genes identified for vB_Pae_BR62a.

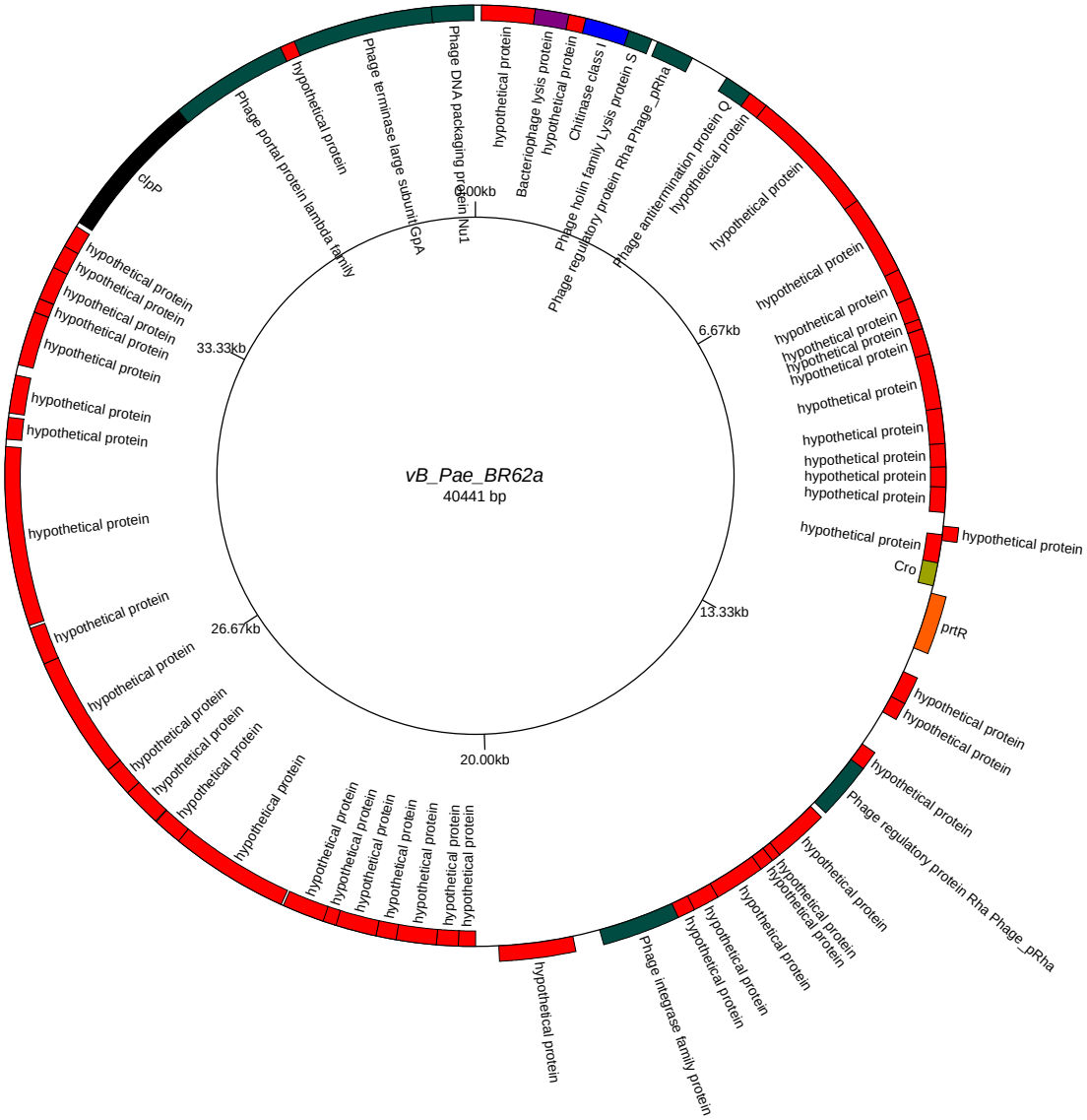
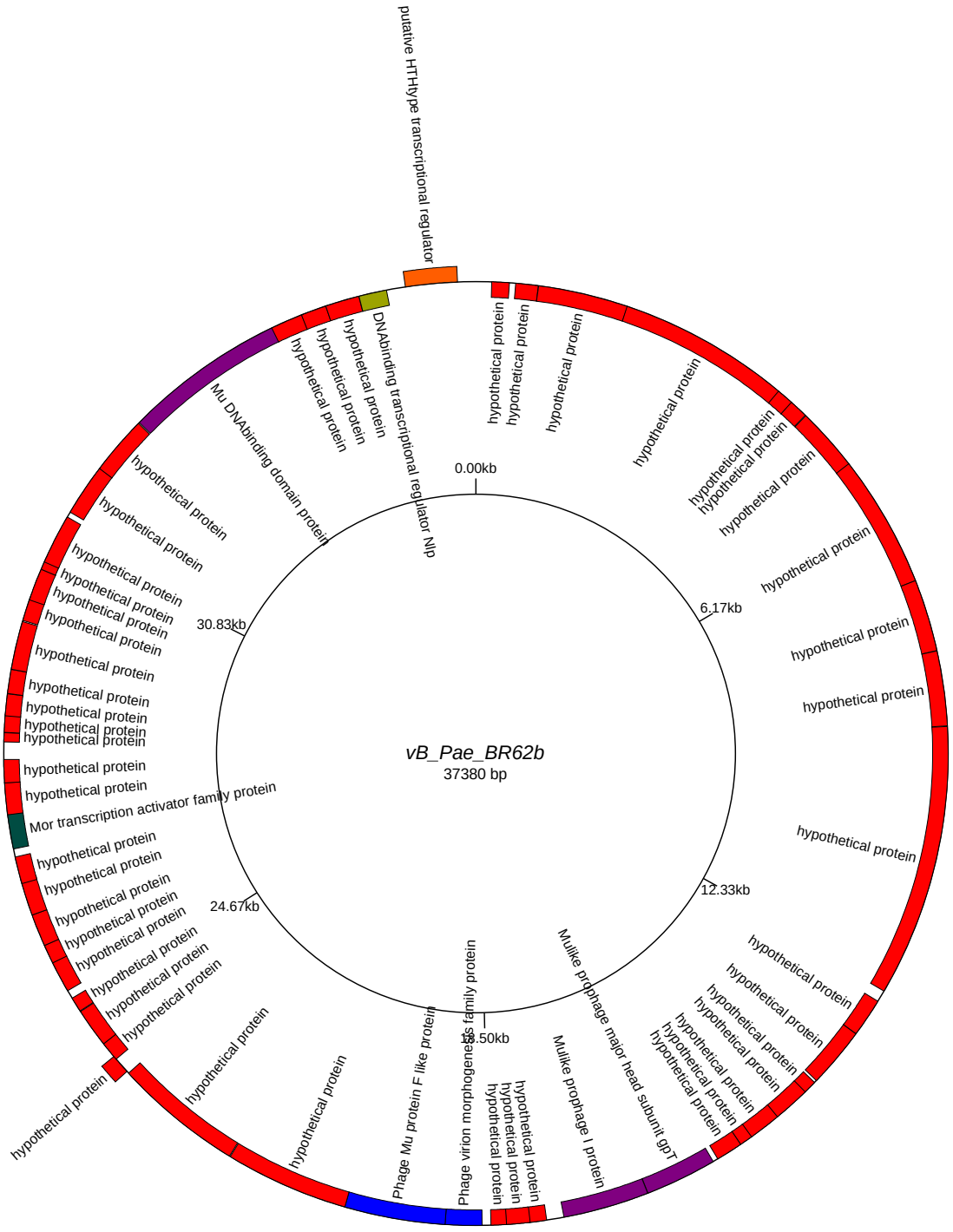


Figure 77: Genome map of vB_Pae_BR62a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR62b genome size was 37.3 kb, the phage was assembled using SPAdes and has a mean coverage of 350. The genome map of this phage is shown in Figure 78. The total number of genes identified was 56 of which 8 were identified with putative functions these can be seen in Table 78. Blastn on viruses only database showed a 27.1 kb / 27.8 kb match with *Pseudomonas* phage JBD24 (GenBank: NC_020203.1). The phage vB_Pae_BR62b had a 64.1 % G-C content. No tRNA gene was identified however; several genes associated with structural proteins were identified.

Start	End	Direction	Putative functional protein
16473	15559	Reverse	Mu-like prophage major head subunit gpT
17574	16477	Reverse	Mu-like prophage I protein
19079	18612	Reverse	Phage virion morphogenesis family protein
20365	19079	Reverse	Phage Mu protein F like protein
27272	26823	Reverse	Mor transcription activator family protein
34728	32659	Reverse	Mu DNA-binding domain protein
36247	35891	Reverse	DNA-binding transcriptional regulator Nlp
36488	37147	Forward	putative HTH-type transcriptional regulator

Table 78: Functional genes identified for vB_Pae_BR62b.


putative HTHtype transcriptional regulator
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
DNAbinding transcriptional regulator NIp
hypothetical protein
hypothetical protein
Mu DNAbinding domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Mor transcription activator family protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Pae_BR63a genome size was 37.7 kb, the phage was assembled using SPAdes and has a mean coverage of 823. The genome map of this phage is shown in Figure 79. The total number of genes identified was 59 of which 10 were identified with putative functions these can be seen in Table 79. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_BR63a had a 64.3 % G-C content. No tRNA gene was identified however; Mu-like phage structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
1223	510	Reverse	HTH-type transcriptional regulator PrtR
1393	1749	Forward	DNA-binding transcriptional regulator Nlp
2451	4445	Forward	Mu DNA-binding domain protein
6658	7176	Forward	Bacteriophage Mu Gam like protein
9980	10429	Forward	Mor transcription activator family protein
16898	18184	Forward	Phage Mu protein F like protein
18184	18651	Forward	Phage virion morphogenesis family protein
19670	20776	Forward	Mu-like prophage I protein
21212	22120	Forward	Mu-like prophage major head subunit gpT
25152	28712	Forward	Prophage tail length tape measure protein

Table 78: Functional genes identified for vB_Pae_BR63a.

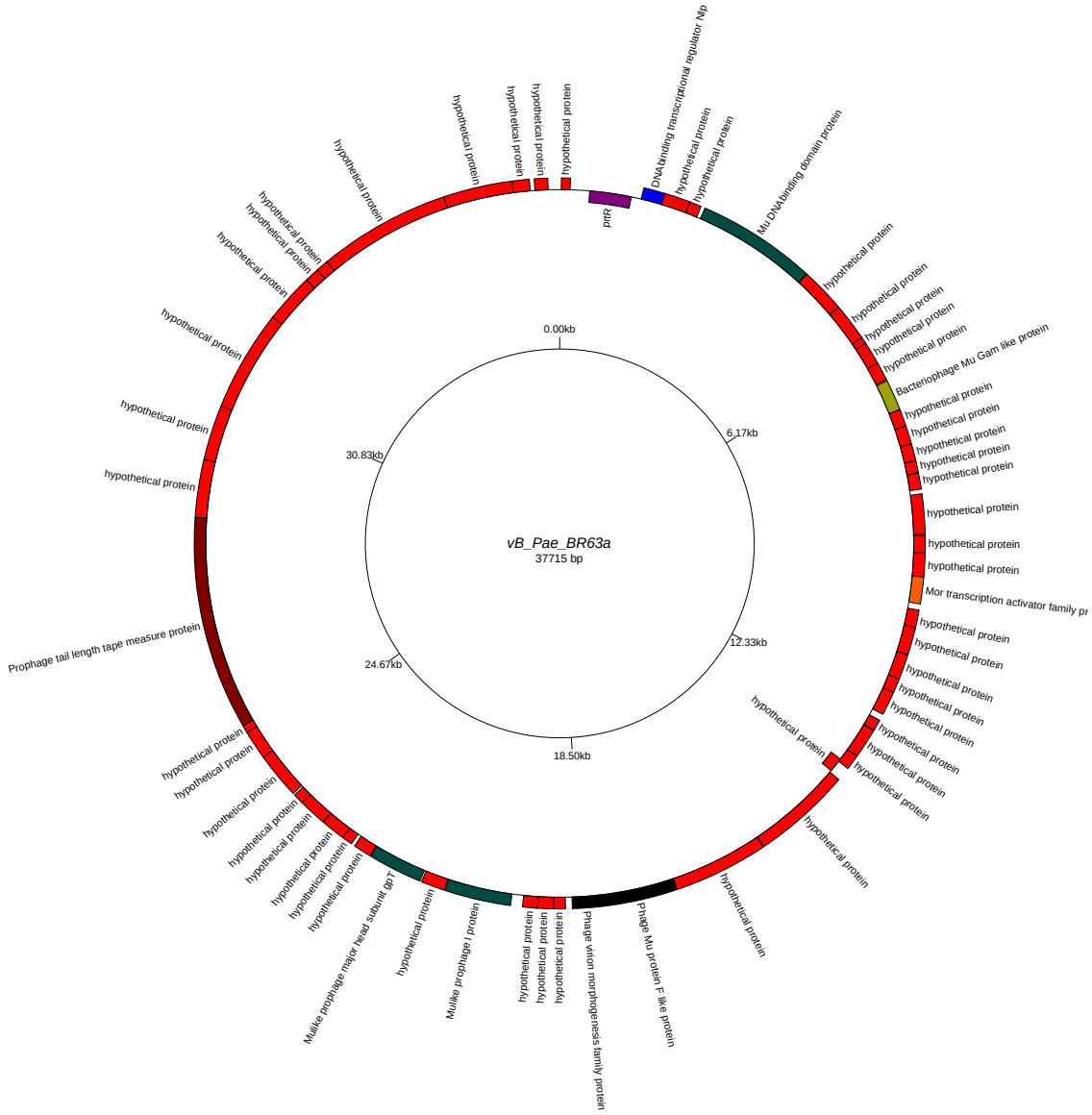


Figure 79: Genome map of vB_Pae_BR63a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR64a genome size was 61.7 kb, the phage was assembled using SPAdes and has a mean coverage of 407. The genome map of this phage is shown in Figure 80. The total number of genes identified was 64 of which 14 were identified with putative functions these can be seen in Table 80. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66 (GenBank: KC262634.1). The phage vB_Pae_BR64a had a 63.7 % G-C content. No tRNA gene was identified however; a CPS-53 integrase like gene was identified.

Start	End	Direction	Putative functional protein
16458	16997	Forward	Phage lysozyme
20063	21298	Forward	Putative prophage CPS-53 integrase
21532	21299	Reverse	Response regulator inhibitor for tor operon
25869	24061	Reverse	C-5 cytosine-specific DNA methylase
28496	25866	Reverse	DNA methylase
28850	28641	Reverse	LexA repressor
30067	28847	Reverse	Recombination-associated protein RdgC
30591	30094	Reverse	Single-stranded DNA-binding protein
36152	35781	Reverse	Carbon storage regulator homolog
38368	37571	Reverse	HTH-type transcriptional regulator PrtR
38476	38676	Forward	Cro
41082	41696	Forward	Bacteriophage Lambda NinG protein
42880	44163	Forward	Phage terminase large subunit
49679	50551	Forward	RyR domain protein

Table 80: Functional genes identified for vB_Pae_BR64a.

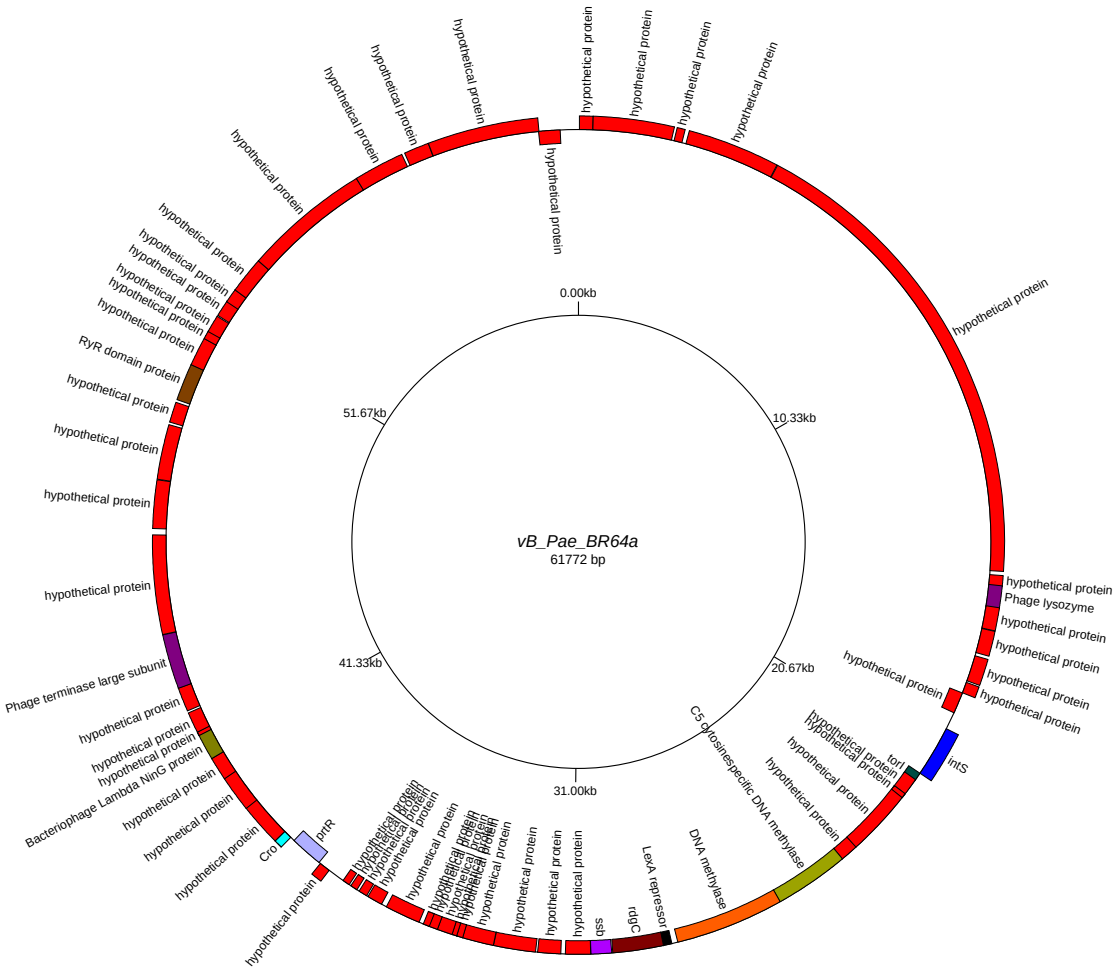


Figure 80: Genome map of vB_Pae_BR64a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR65a genome size was 40.4 kb, the phage was assembled using SPAdes and extended using PriceTI. The genome map of this phage is shown in Figure 81. The total number of genes identified was 61 of which 13 were identified with putative functions these can be seen in Table 81. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR65a had a 61.5 % G-C content. No tRNA gene was identified however; a *Cro* gene and phage integrase like gene was identified.

Start	**End**	**Direction**	**Putative functional protein**
1346	876	Reverse	Bacteriophage lysis protein
2199	1582	Reverse	Chitinase class I
2528	2196	Reverse	Phage holin family (Lysis protein S)
3140	2613	Reverse	Phage regulatory protein Rha (Phage_pRha)
4068	3679	Reverse	Phage antitermination protein Q
11684	11364	Reverse	Cro
11787	12575	Forward	HTH-type transcriptional regulator PrtR
14329	15096	Forward	Phage regulatory protein Rha (Phage_pRha)
17505	18557	Forward	Phage integrase family protein
36129	34048	Reverse	ATP-dependent Clp protease proteolytic subunit
37747	36101	Reverse	Phage portal protein, lambda family
39917	37953	Reverse	Phage terminase large subunit (GpA)
40230	39889	Reverse	Phage DNA packaging protein Nu1

Table 81: Functional genes identified for vB_Pae_BR65a.

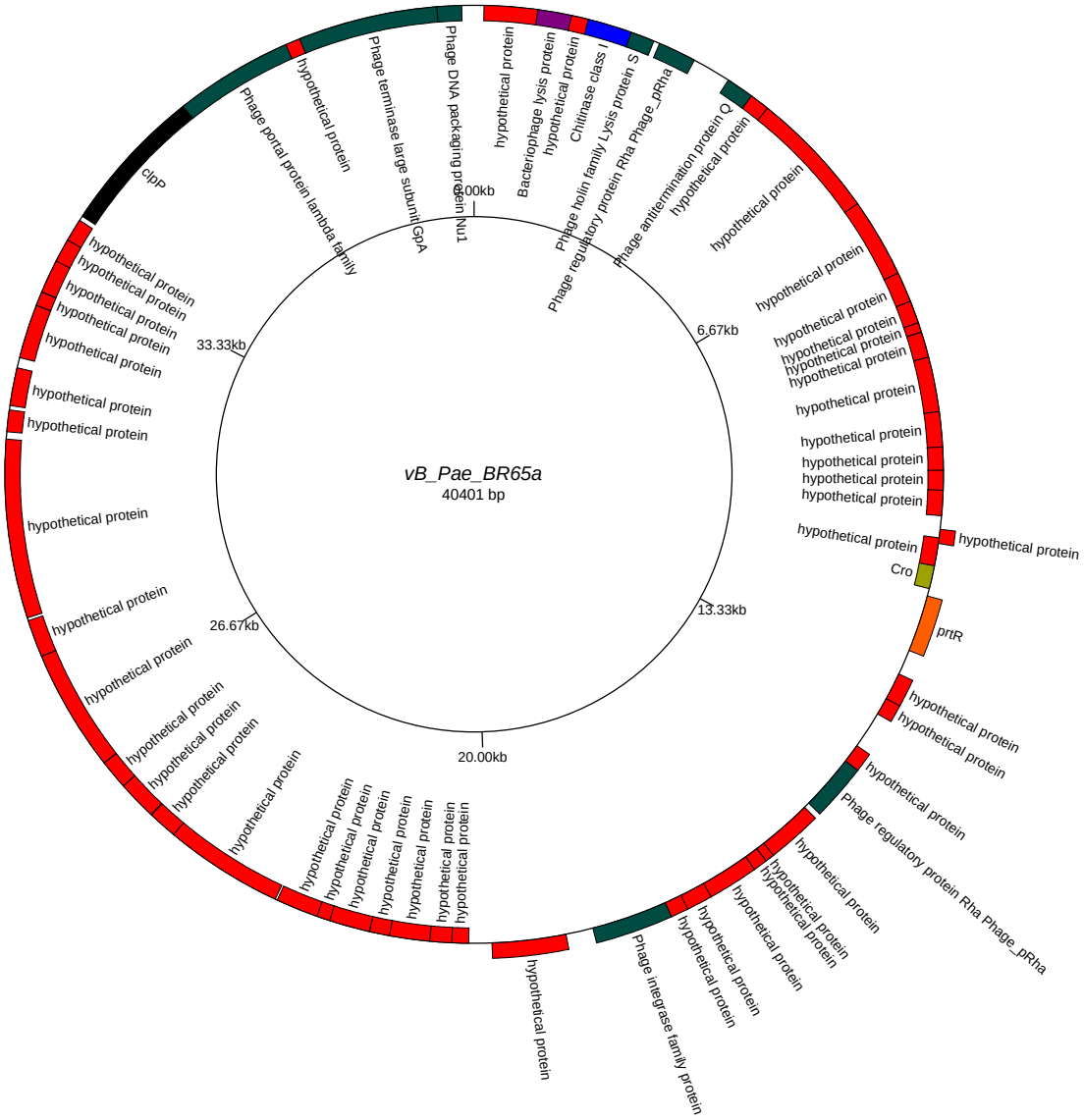
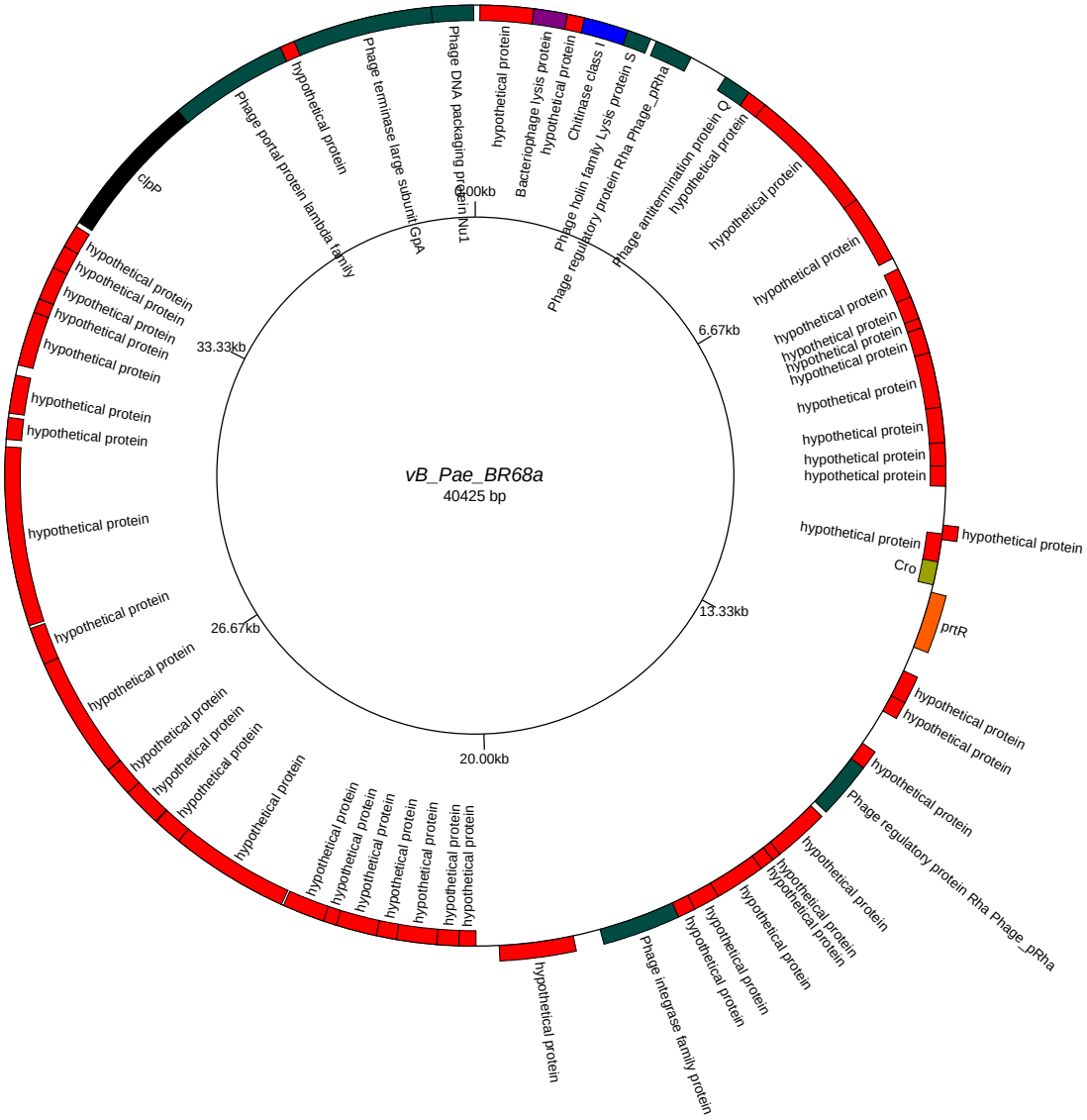


Figure 81: Genome map of vB_Pae_BR65a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR68a genome size was 40.4 kb, the phage was assembled using SPAdes and extended using PriceTI. The genome map of this phage is shown in Figure 82. The total number of genes identified was 60 of which 13 were identified with putative functions these can be seen in Table 82. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR68a had a 61.5 % G-C content. No tRNA gene was identified however; a *Cro* gene and phage integrase like gene was identified.

Start	End	Direction	Putative functional protein
1274	804	Reverse	Bacteriophage lysis protein
2127	1510	Reverse	Chitinase class I
2456	2124	Reverse	Phage holin family (Lysis protein S)
3068	2541	Reverse	Phage regulatory protein Rha (Phage_pRha)
3996	3607	Reverse	Phage antitermination protein Q
11614	11294	Reverse	Cro
11717	12505	Forward	HTH-type transcriptional regulator PrtR
14259	15026	Forward	Phage regulatory protein Rha (Phage_pRha)
17435	18487	Forward	Phage integrase family protein
36058	33977	Reverse	ATP-dependent Clp protease proteolytic subunit
37676	36030	Reverse	Phage portal protein, lambda family
39846	37882	Reverse	Phage terminase large subunit (GpA)
40402	39818	Reverse	Phage DNA packaging protein Nu1

Table 82: Functional genes identified for vB_Pae_BR68a.


vB_Pae_BR68a
40425 bp
0.00kb
6.67kb
13.33kb
20.00kb
26.67kb
33.33kb
Phage DNA packaging protein Nu1
hypothetical protein
Bacteriophage lysis protein
hypothetical protein
Chitinase class I
Lysis protein S
Phage holin family
Phage regulatory protein Rha Phage_pRha
Phage antitermination protein Q
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Cro
prtR
hypothetical protein
hypothetical protein
hypothetical protein
Phage regulatory protein Rha Phage_pRha
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage integrase family protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Pae_BR69a genome size was 46.4 kb, the phage was assembled using SPAdes and has a mean coverage of 236. The genome map of this phage is shown in Figure 83. The total number of genes identified was 72 of which 15 were identified with putative functions these can be seen in Table 83. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_BR69a had a 59.3 % G-C content. A tRNA gene was identified; CII and Ig-like domains were identified along with recombinase genes.

Start	End	Direction	Putative functional protein
43	1107	Forward	IgA-specific serine endopeptidase autotransporter precursor
1115	1861	Forward	RecT family protein
1845	2465	Forward	YqaJ-like viral recombinase domain protein
2462	2797	Forward	LytTr DNA-binding domain protein
10559	9549	Forward	site-specific tyrosine recombinase XerC
10720	10633	Reverse	tRNA-Ser(cga)
12224	11790	Reverse	Lysozyme RrrD
12827	13726	Forward	BRO family, N-terminal domain
15867	14221	Reverse	D-glucuronyl C5-epimerase C-terminus
23322	22762	Reverse	AP2 domain protein
28403	27408	Reverse	Bacterial Ig-like domain (group 2)
35018	33288	Reverse	Phage Mu protein F like protein
37669	36551	Reverse	Phage terminase large subunit
42494	41847	Reverse	Bacteriophage Lambda NinG protein
46293	44908	Reverse	Replicative DNA helicase

Table 83: Functional genes identified for vB_Pae_BR69a.

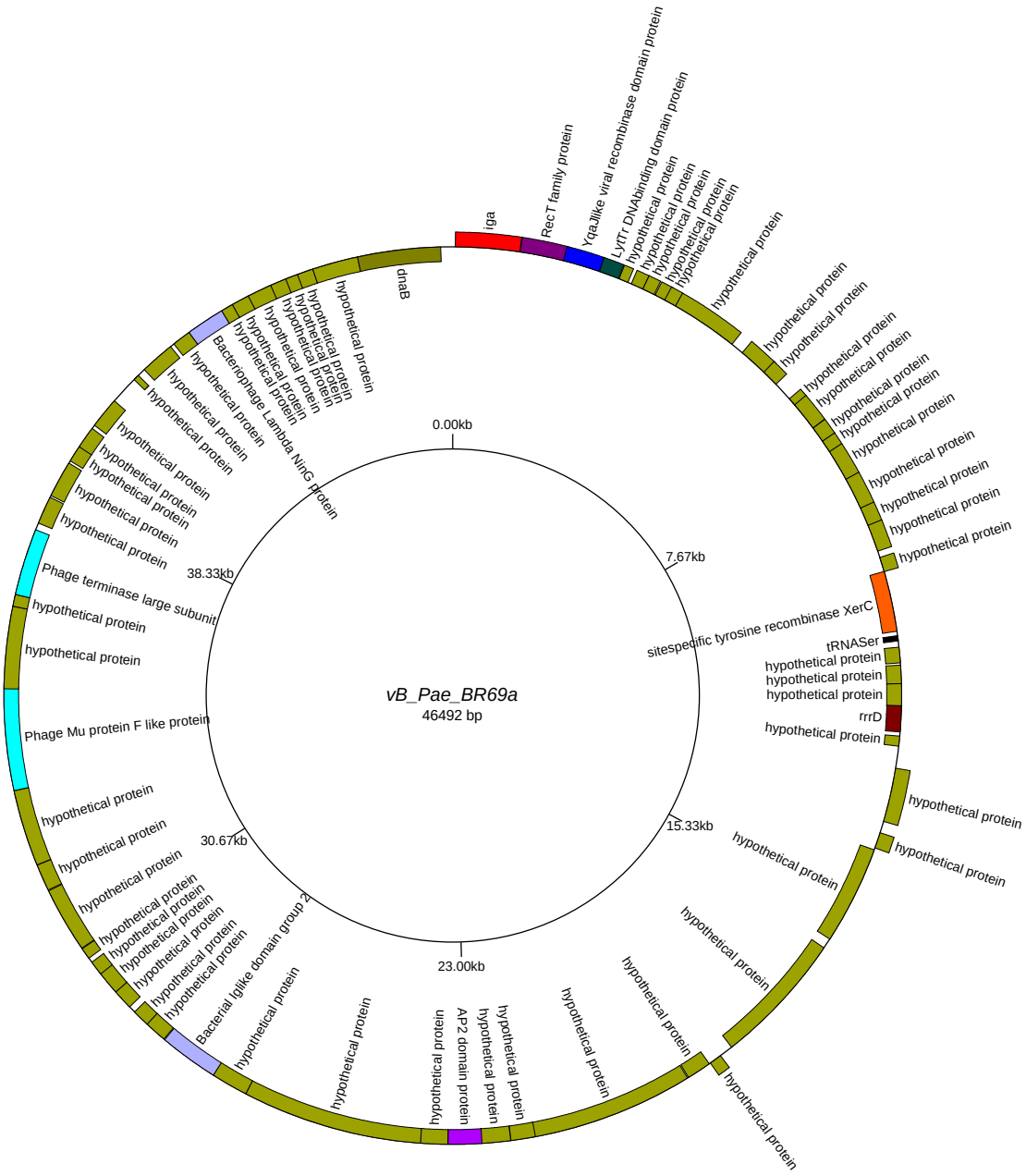


Figure 83: Genome map of vB_Pae_BR69a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR69b genome size was 38.2 kb, the phage was assembled using SPAdes and has a mean coverage of 345. The genome map of this phage is shown in Figure 84. The total number of genes identified was 60 of which 10 were identified with putative functions these can be seen in Table 84. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_BR63b had a 64.2 % G-C content. No tRNA gene was identified however; Mu-like phage structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
13129	9569	Reverse	Prophage tail length tape measure protein
17069	16161	Reverse	Mu-like prophage major head subunit gpT
18611	17505	Reverse	Mu-like prophage I protein
20097	19630	Reverse	Phage virion morphogenesis family protein
21383	20097	Reverse	Phage Mu protein F like protein
28301	27852	Reverse	Mor transcription activator family protein
31623	31105	Reverse	Bacteriophage Mu Gam like protein
35830	33836	Reverse	Mu DNA-binding domain protein
36888	36532	Reverse	DNA-binding transcriptional regulator Nlp
37058	37771	Forward	HTH-type transcriptional regulator PrtR

Table 84: Functional genes identified for vB_Pae_BR69b.

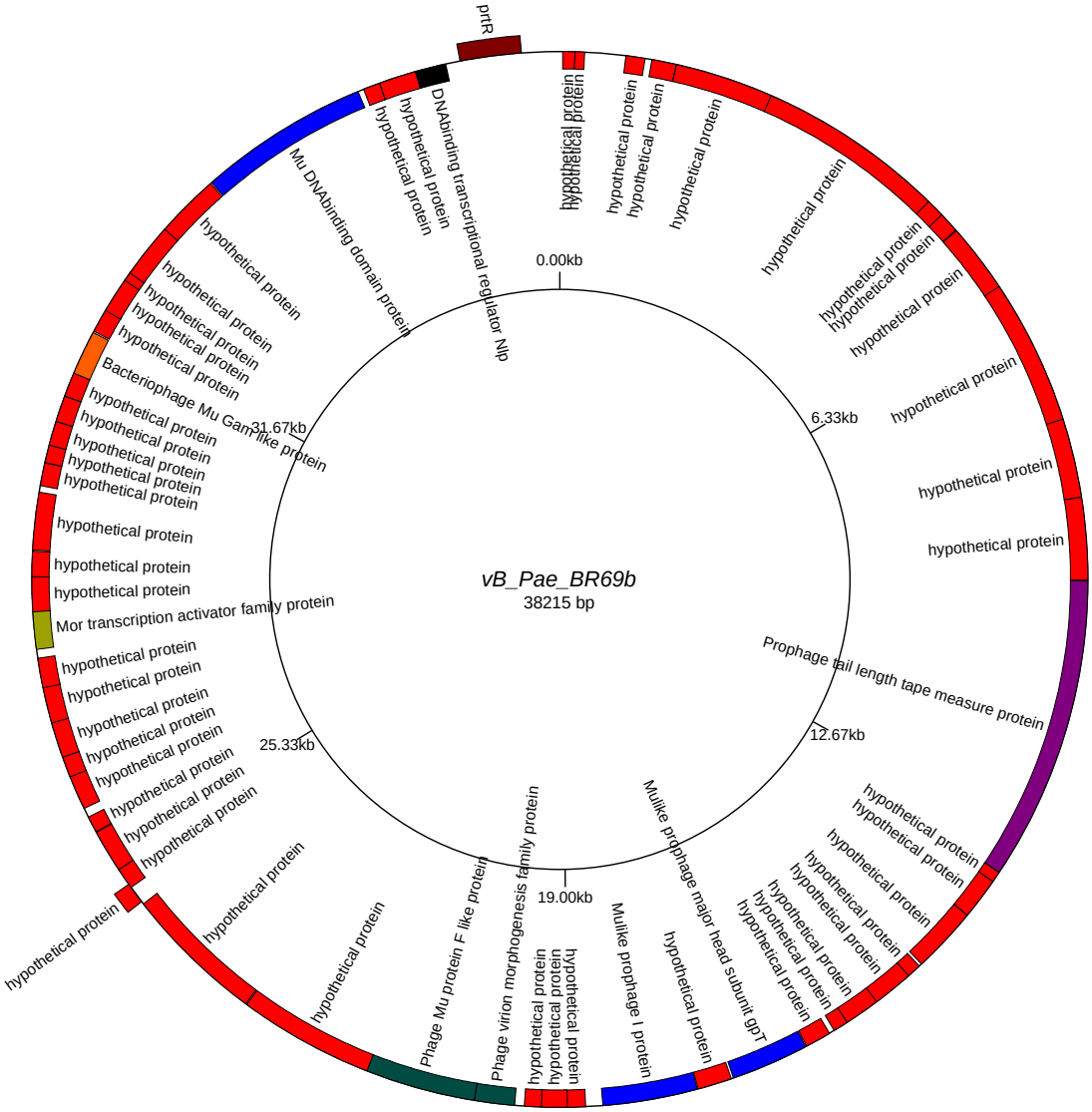


Figure 84: Genome map of vB_Pae_BR69b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR72a genome size was 37.4 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 85. The total number of genes identified was 56 of which 8 were identified with putative functions these can be seen in Table 85. Blastn on viruses only database showed a 27.1 kb / 27.8 kb match with *Pseudomonas* phage JBD24 (GenBank: NC_020203.1). The phage vB_Pae_BR72a had a 64.1 % G-C content. No tRNA gene was identified however; several genes associated with structural proteins were identified.

Start	**End**	**Direction**	**Putative functional proteins**
16546	15632	Reverse	Mu-like prophage major head subunit gpT
17647	16550	Reverse	Mu-like prophage I protein
19152	18685	Reverse	Phage virion morphogenesis family protein
20438	19152	Reverse	Phage Mu protein F like protein
27345	26896	Reverse	Mor transcription activator family protein
34801	32732	Reverse	Mu DNA-binding domain protein
36320	35964	Reverse	DNA-binding transcriptional regulator Nlp
36561	37220	Forward	putative HTH-type transcriptional regulator

Table 85: Functional genes identified for vB_Pae_BR72a.

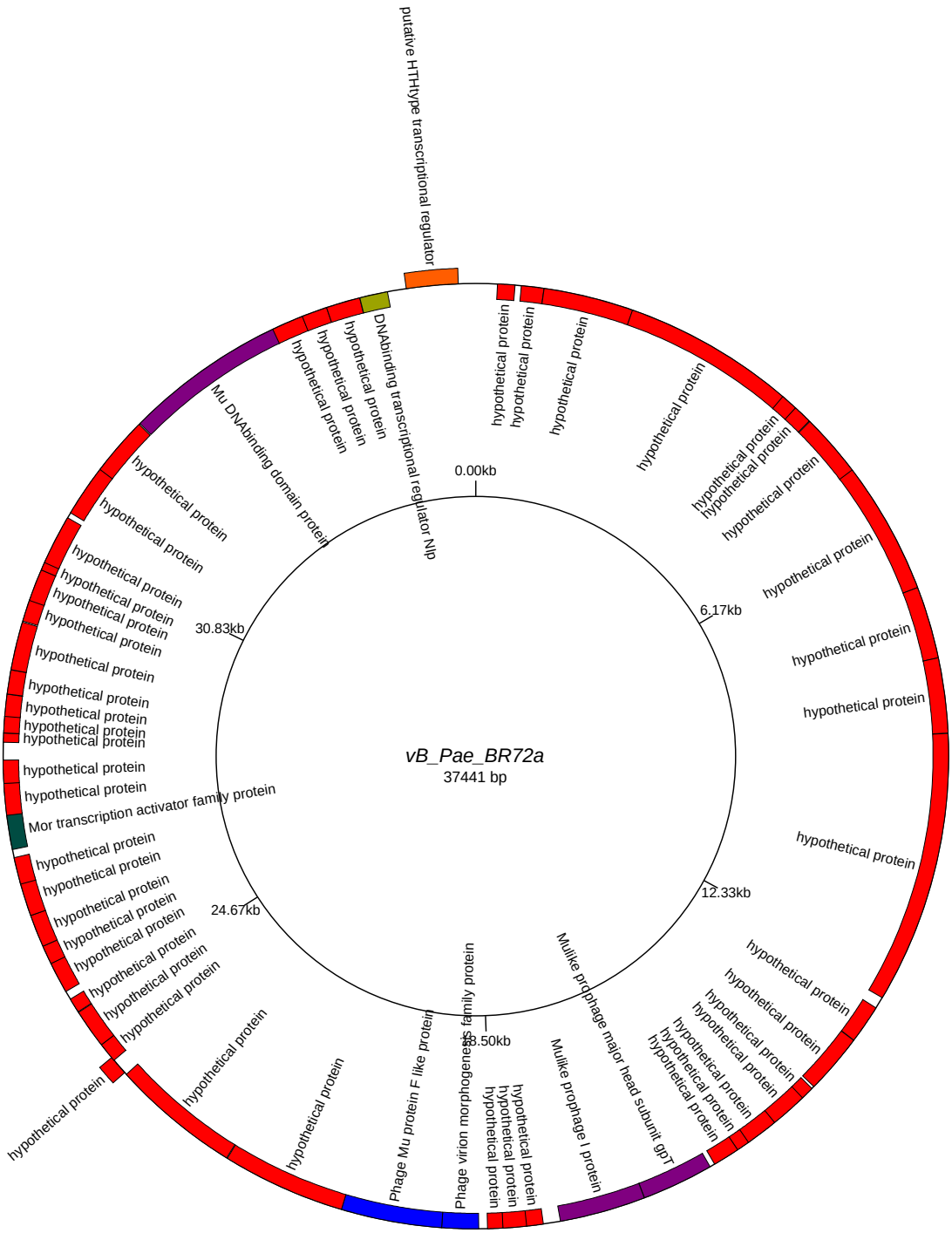
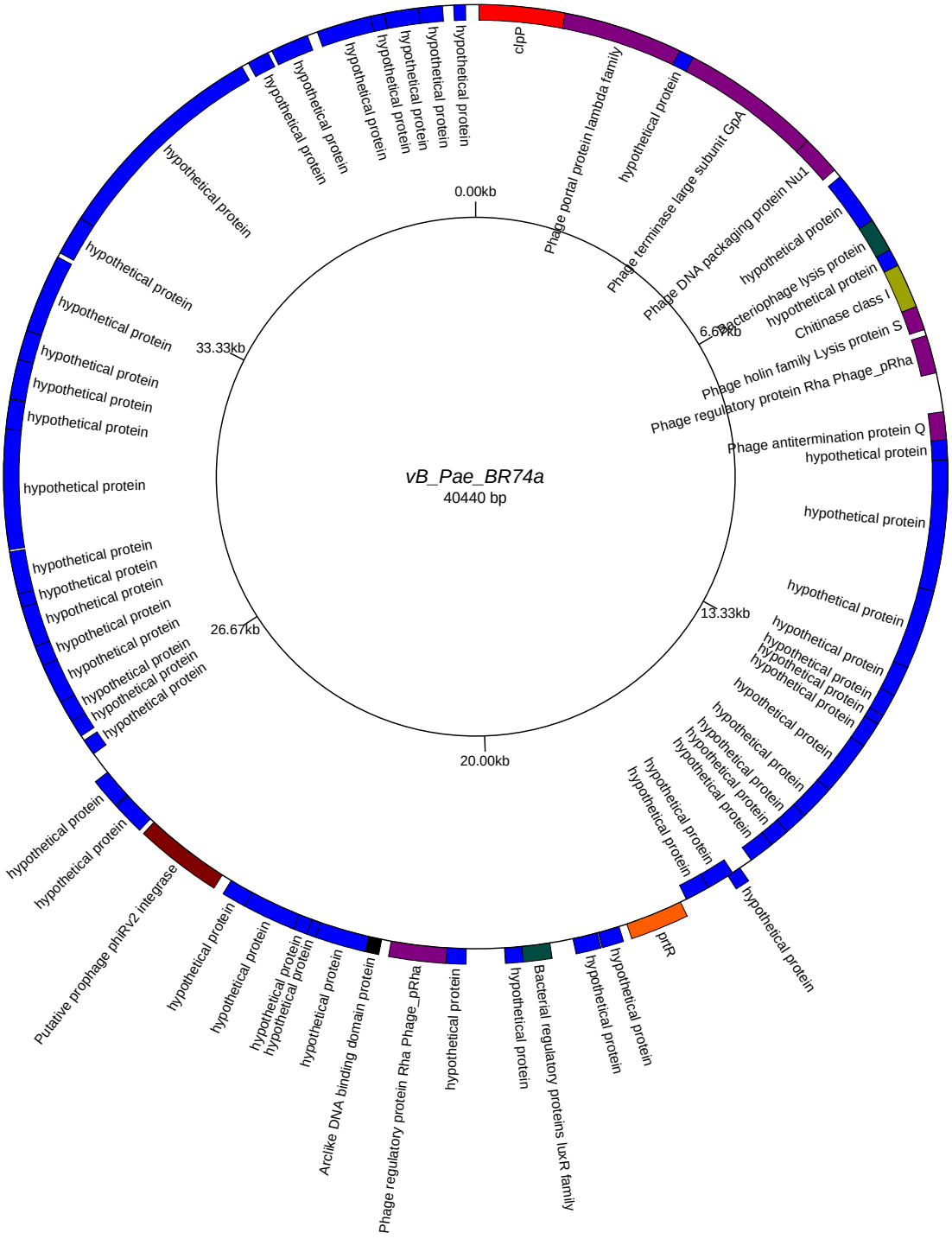


Figure 85: Genome map of vB_Pae_BR72a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR74a genome size was 40.4 kb, the phage was assembled using SPAdes and had a mean coverage of 793. The genome map of this phage is shown in Figure 86. The total number of genes identified was 65 of which 14 were identified with putative functions these can be seen in Table 86. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR74a had a 61.4 % G-C content. No tRNA gene was identified however; a *Cro* gene and phage integrase like gene was identified.

Start	End	Direction	Putative functional protein
1255	44	Reverse	ATP-dependent Clp protease proteolytic subunit
2873	1227	Reverse	Phage portal protein, lambda family
5043	3079	Reverse	Phage terminase large subunit (GpA)
5560	5015	Reverse	Phage DNA packaging protein Nu1
6887	6417	Reverse	Bacteriophage lysis protein
7740	7123	Reverse	Chitinase class I
8069	7737	Reverse	Phage holin family (Lysis protein S)
8681	8154	Reverse	Phage regulatory protein Rha (Phage_pRha)
9609	9220	Reverse	Phage antitermination protein Q
17332	18123	Reverse	HTH-type transcriptional regulator PrtR
19206	19592	Forward	Bacterial regulatory proteins, luxR family
20611	21378	Forward	Phage regulatory protein Rha (Phage_pRha)
21509	21688	Forward	Arc-like DNA binding domain protein
23868	25046	Forward	Putative prophage phiRv2 integrase

Table 86: Functional genes identified for vB_Pae_BR74a.


vB_Pae_BR74a
40440 bp
0.00kb
6.67kb
13.33kb
20.00kb
26.67kb
33.33kb
clpp
Phage portal protein lambda family
hypothetical protein
Phage terminase large subunit GpA
Phage DNA packaging protein Nu1
hypothetical protein
Bacteriophage lysis protein
hypothetical protein
Chitinase class I
Phage holin family Lysis protein S
Phage regulatory protein Rha Phage_pRha
Phage antitermination protein Q
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Pae_BR75a genome size was 61.7 kb, the phage was assembled using SPAdes and has a mean coverage of 33. The genome map of this phage is shown in Figure 87. The total number of genes identified was 66 of which 13 were identified with putative functions these can be seen in Table 87. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66 (GenBank: KC262634.1). The phage vB_Pae_BR75a had a 63.7 % G-C content. No tRNA gene was identified however; a CPS-53 integrase like gene was identified.

Start	End	Direction	Putative functional protein
9147	10382	Forward	Putative prophage CPS-53 integrase
10616	10383	Reverse	Response regulator inhibitor for tor operon
14953	13145	Reverse	C-5 cytosine-specific DNA methylase
17580	14950	Reverse	DNA methylase
17934	17725	Forward	LexA repressor
19151	17931	Reverse	Recombination-associated protein RdgC
19675	19178	Reverse	Single-stranded DNA-binding protein
25236	24865	Reverse	Carbon storage regulator homolog
27452	26655	Reverse	HTH-type transcriptional regulator PrtR
27560	27760	Forward	Cro
30166	30780	Forward	Bacteriophage Lambda NinG protein
31964	33247	Forward	Phage terminase large subunit
38763	39635	Forward	RyR domain protein

Table 87: Functional genes identified for vB_Pae_BR75a.

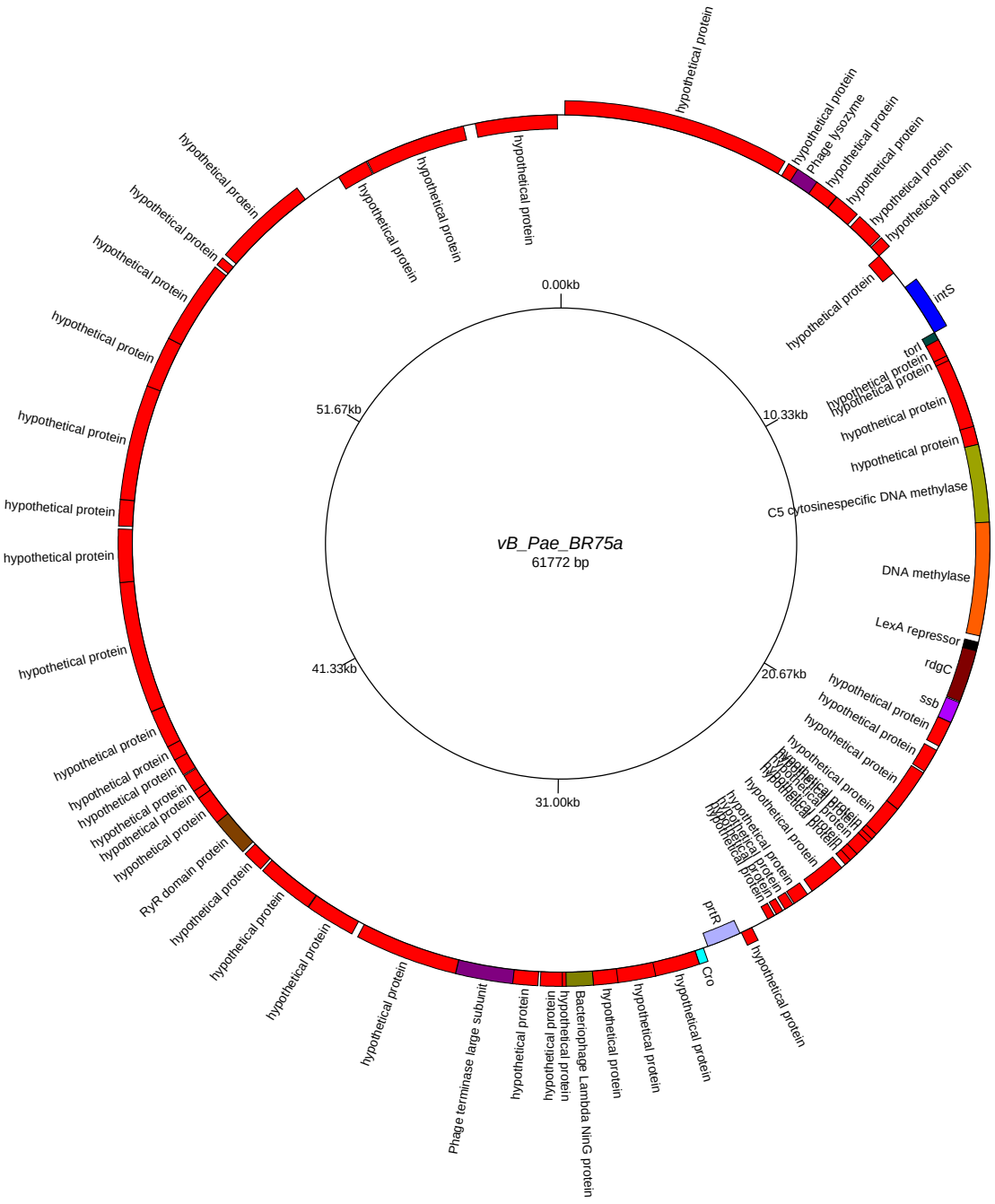


Figure 87: Genome map of vB_Pae_BR75a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR75b genome size was 37.7 kb, the phage was assembled using SPAdes and has a mean coverage of 59. The genome map of this phage is shown in Figure 88. The total number of genes identified was 58 of which 10 were identified with putative functions these can be seen in Table 88. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_BR75b had a 64.3 % G-C content. No tRNA gene was identified however; Mu-like phage structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
1169	456	Reverse	HTH-type transcriptional regulator PrtR
1339	1695	Forward	DNA-binding transcriptional regulator Nlp
2397	4391	Forward	Mu DNA-binding domain protein
6604	7122	Forward	Bacteriophage Mu Gam like protein
9926	10375	Forward	Mor transcription activator family protein
16844	18130	Forward	Phage Mu protein F like protein
18130	18597	Forward	Phage virion morphogenesis family protein
19616	20722	Forward	Mu-like prophage I protein
21158	22066	Forward	Mu-like prophage major head subunit gpT
25098	28658	Forward	Prophage tail length tape measure protein

Table 88: Functional genes identified for vB_Pae_BR75b.

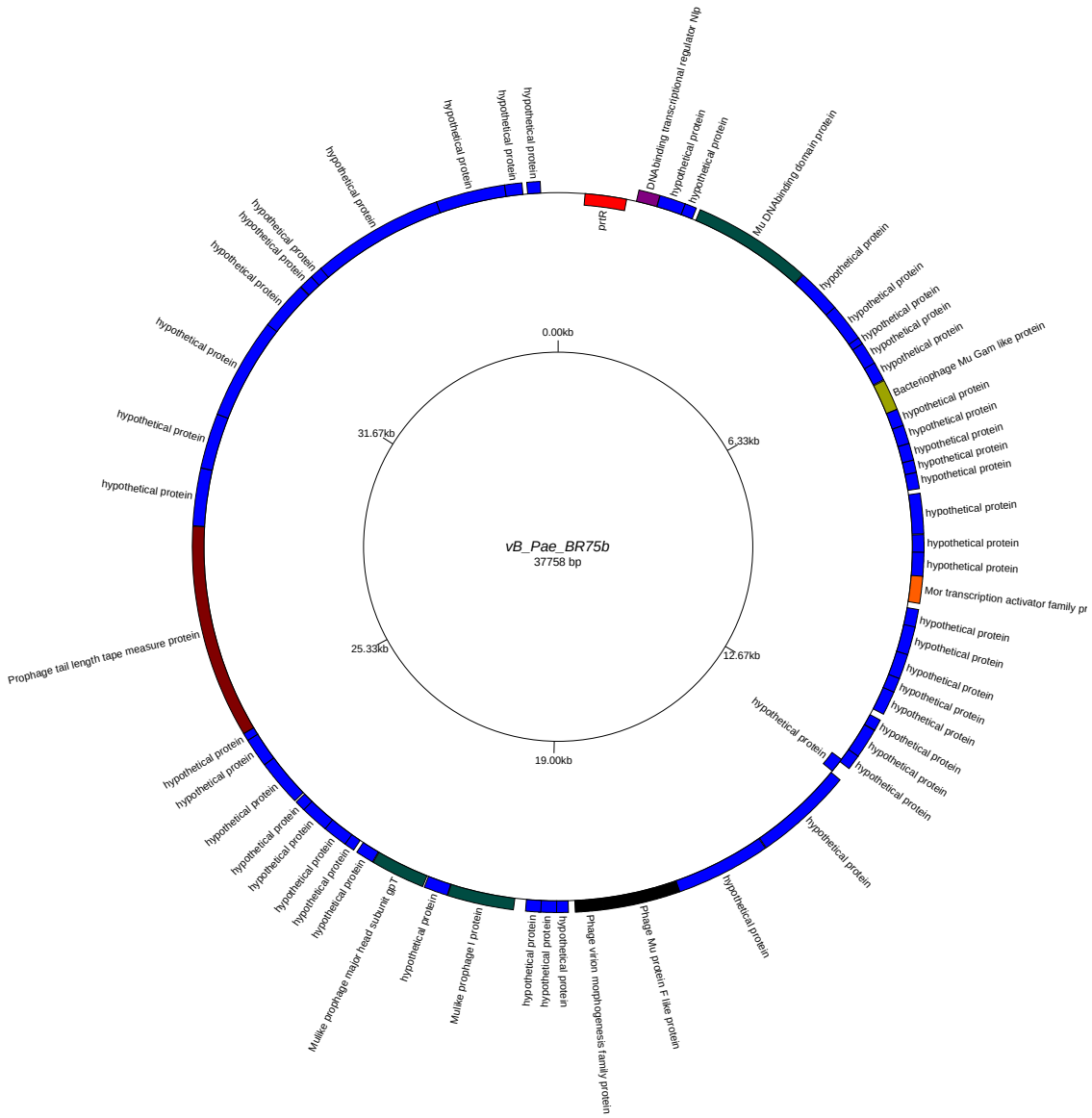
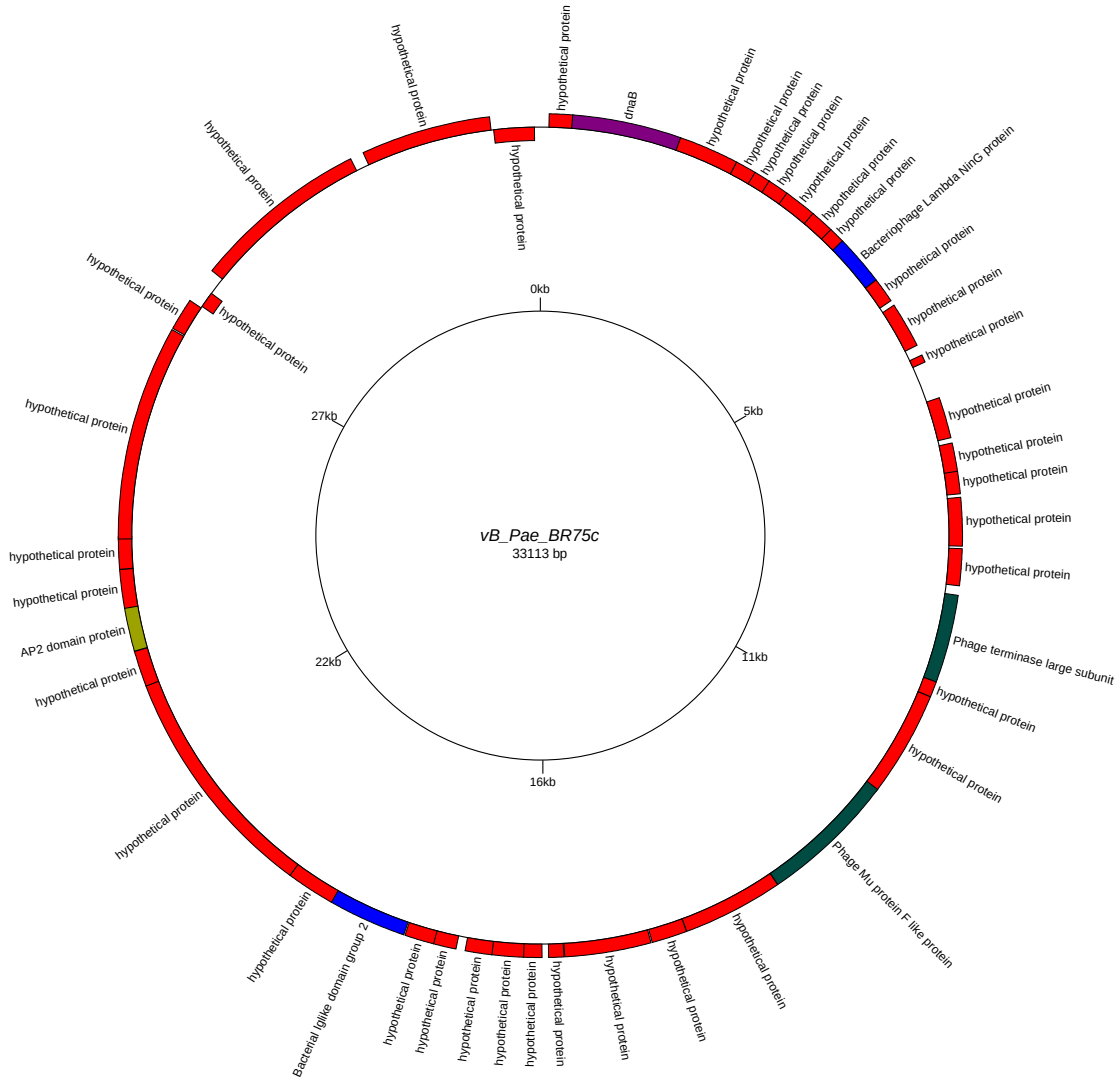


Figure 88: Genome map of vB_Pae_BR75b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR75c genome size was 33.1 kb, the phage was assembled using SPAdes and has a mean coverage of 46. The genome map of this phage is shown in Figure 89. The total number of genes identified was 44 of which 7 were identified with putative functions these can be seen in Table 89. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_BR75c had a 59 % G-C content. No tRNA gene was identified however; Ig-like domain was identified along with recombinase genes.

Start	End	Direction	Putative functional protein
403	1788	Forward	Replicative DNA helicase
4202	4849	Forward	Bacteriophage Lambda NinG protein
9027	10145	Forward	Phage terminase large subunit
11678	13408	Forward	Phage Mu protein F like protein
18293	19288	Forward	Bacterial Ig-like domain (group 2)
23374	23934	Forward	AP2 domain protein
30829	32475	Forward	D-glucuronyl C5-epimerase C-terminus

Table 89: Functional genes identified for vB_Pae_BR75c.


vB_Pae_BR75c
33113 bp
0kb
5kb
11kb
16kb
22kb
27kb
hypothetical protein
dnaB
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Bacteriophage Lambda NinG protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage terminase large subunit
hypothetical protein
hypothetical protein
Phage Mu protein F like protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Pae_BR76a genome size was 35.3 kb, the phage was assembled using SPAdes and has a mean coverage of 22. The genome map of this phage is shown in Figure 90. The total number of genes identified was 43 of which 26 were identified with putative functions these can be seen in Table 90. Blastn on viruses only database showed a 9.4 kb / 9.8 kb match with *Pseudomonas* phage phiCTX (GenBank: NC_003278.1). The phage vB_Pae_BR76a had a 62.4 % G-C content. No tRNA gene was identified however; Ig-like domain, recombinase genes, lysogenic, structural and a CTX gene were identified.

Start	End	Direction	Putative functional protein
1954	785	Reverse	site-specific tyrosine recombinase XerD
3953	2175	Reverse	C-5 cytosine-specific DNA methylase
8535	8242	Reverse	Ogr/Delta-like zinc finger
9351	9767	Forward	Bacteriophage CI repressor helix-turn-helix domain protein
12977	11703	Reverse	Phage late control gene D protein (GPD)
13414	12974	Reverse	Phage P2 GpU
16179	13420	Reverse	Phage-related minor tail protein
16288	16169	Reverse	Phage P2 GpE
16635	16297	Reverse	Phage tail protein E
17204	16689	Reverse	Phage tail tube protein FII
18436	17261	Reverse	Phage tail sheath protein
21808	21272	Reverse	Phage tail protein (Tail_P2_I)
22722	21808	Reverse	Baseplate J-like protein
23063	22719	Reverse	Gene 25-like lysozyme
23641	23060	Reverse	Phage-related baseplate assembly protein
25091	24633	Reverse	Phage virion morphogenesis family protein
25620	25084	Reverse	P2 phage tail completion protein R (GpR)
26962	26156	Reverse	Zinc D-Ala-D-Ala carboxypeptidase precursor
27823	27611	Reverse	Phage Tail Protein X
28284	27823	Reverse	Phage head completion protein (GPL)
29089	28388	Reverse	Phage small terminase subunit

30111	29095	Reverse	Phage major capsid protein, P2 family
30968	30147	Reverse	Phage capsid scaffolding protein (GPO) serine peptidase
31103	32887	Reverse	Terminase-like family protein
32884	33936	Forward	Phage portal protein
34991	34131	Reverse	Cytotoxin precursor

Table 90: Functional genes identified for vB_Pae_BR76a.

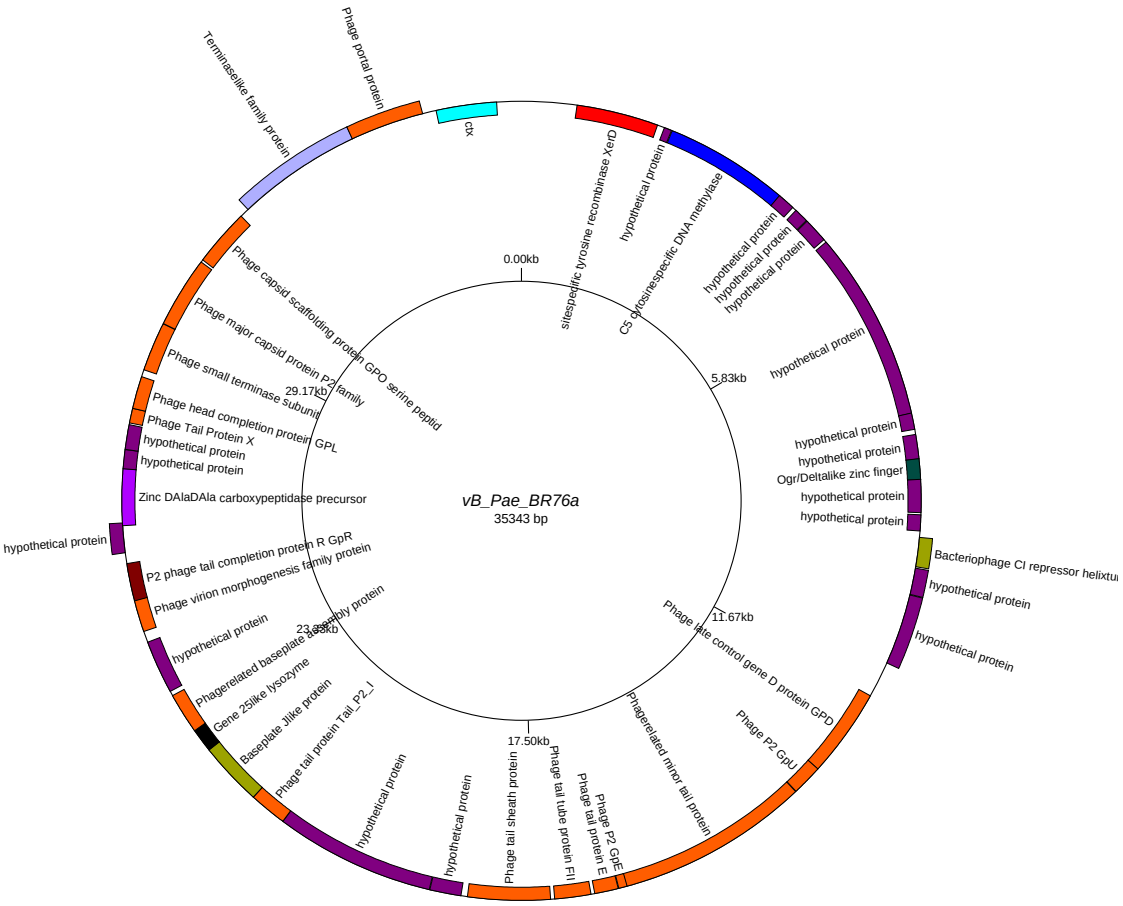
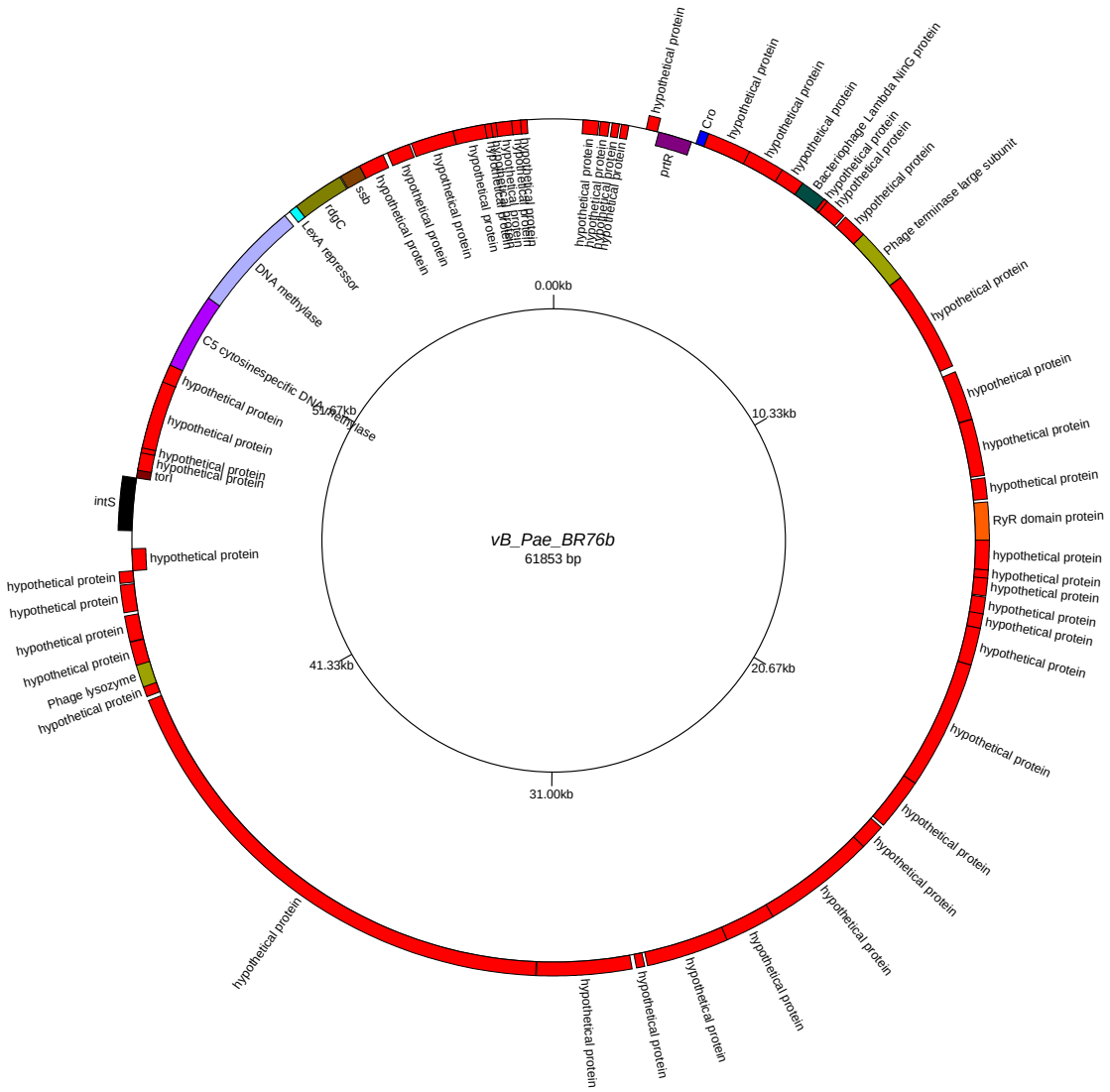


Figure 90: Genome map of vB_Pae_BR76a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR76b genome size was 61.8 kb, the phage was assembled using SPAdes and has a mean coverage of 15. The genome map of this phage is shown in Figure 91. The total number of genes identified was 62 of which 13 were identified with putative functions these can be seen in Table 91. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66 (GenBank: KC262634.1). The phage vB_Pae_BR76b had a 63.7 % G-C content. No tRNA gene was identified however; a CPS-53 integrase like gene, lysogenic and regulatory genes were identified.

Start	End	Direction	Putative functional protein
1061	690	Reverse	Carbon storage regulator homolog
3277	2480	Reverse	HTH-type transcriptional regulator PrtR
3385	3585	Forward	Cro
5991	6605	Forward	Bacteriophage Lambda NinG protein
7789	9072	Forward	Phage terminase large subunit
14588	15460	Forward	RyR domain protein
43012	43551	Forward	Phage lysozyme
46617	47852	Forward	Putative prophage CPS-53 integrase
48086	47853	Reverse	Response regulator inhibitor for tor operon
52423	50615	Reverse	C-5 cytosine-specific DNA methylase
55050	52420	Reverse	DNA methylase
55404	55195	Reverse	LexA repressor
56621	55401	Reverse	Recombination-associated protein RdgC
57145	56648	Reverse	Single-stranded DNA-binding protein

Table 91: Functional genes identified for vB_Pae_BR76b.


vB_Pae_BR76b
61853 bp
0.00kb
10.33kb
20.67kb
31.00kb
41.33kb
51.67kb
hypothetical protein
Cro
hypothetical protein
hypothetical protein
Bacteriophage Lambda NinG protein
hypothetical protein
hypothetical protein
Phage terminase large subunit
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
RyR domain protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Pae_BR77a genome size was 61.7 kb, the phage was assembled using SPAdes and has a mean coverage of 413. The genome map of this phage is shown in Figure 92. The total number of genes identified was 62 of which 14 were identified with putative functions these can be seen in Table 92. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66 (GenBank: KC262634.1). The phage vB_Pae_BR77a had a 63.7 % G-C content. No tRNA gene was identified however; a CPS-53 integrase like gene was identified.

Start	End	Direction	Putative functional protein
4533	5030	Forward	Single-stranded DNA-binding protein
5057	6277	Forward	Recombination-associated protein RdgC
6274	6483	Forward	LexA repressor
6628	9258	Forward	DNA methylase
9255	11063	Forward	C-5 cytosine-specific DNA methylase
13592	13825	Forward	Response regulator inhibitor for tor operon
15061	13826	Reverse	Putative prophage CPS-53 integrase
18666	18127	Reverse	Phage lysozyme
47090	46218	Reverse	RyR domain protein
53889	52606	Reverse	Phage terminase large subunit
55687	55073	Reverse	Bacteriophage Lambda NinG protein
58293	58093	Reverse	Cro
58401	59198	Forward	HTH-type transcriptional regulator PrtR
60617	60988	Forward	Carbon storage regulator homolog

Table 92: Functional genes identified for vB_Pae_BR77a.

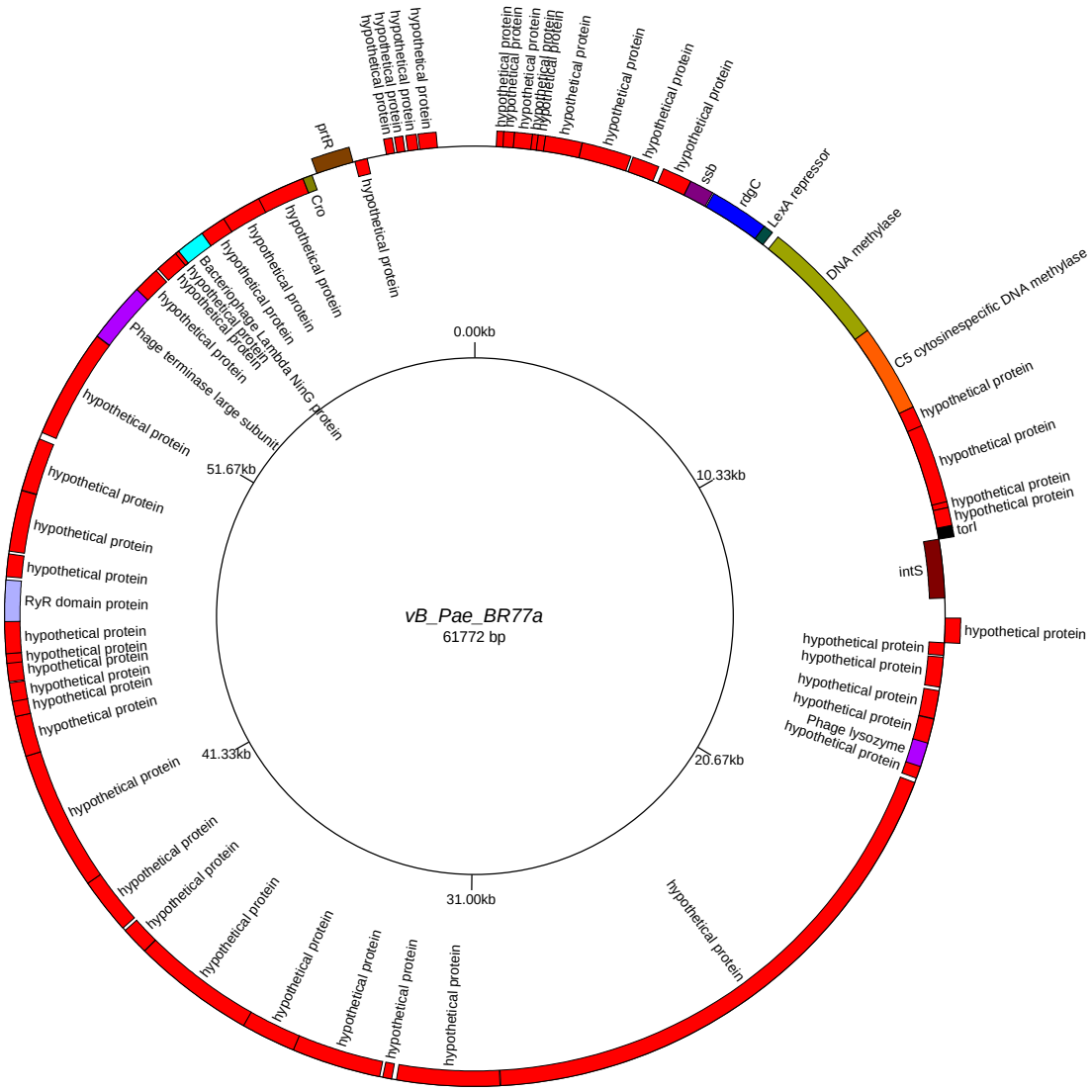


Figure 92: Genome map of vB_Pae_BR77a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR78a genome size was 43.1 kb, the phage was assembled using SPAdes and extended using PriceTI. The genome map of this phage is shown in Figure 93. The total number of genes identified was 67 of which 15 were identified with putative functions these can be seen in Table 93. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR78a had a 61.3 % G-C content. No tRNA gene was identified however; a *Cro* gene and phage integrase like gene was identified.

Start	End	Direction	Putative functional protein
367	1083	Forward	HTH-type transcriptional regulator PrtR
6433	6822	Forward	Phage antitermination protein Q
7361	7888	Forward	Phage regulatory protein Rha (Phage_pRha)
7973	8305	Forward	Phage holin family (Lysis protein S)
8302	8919	Forward	Chitinase class I
9155	9625	Forward	Bacteriophage lysis protein
10482	11027	Forward	Phage DNA packaging protein Nu1
10999	12963	Forward	Phage terminase large subunit (GpA)
13169	14815	Forward	Phage portal protein, lambda family
14787	16868	Forward	ATP-dependent Clp protease proteolytic subunit
33362	32184	Reverse	Putative prophage phiRv2 integrase
35721	35542	Reverse	Arc-like DNA binding domain protein
36619	35852	Reverse	Phage regulatory protein Rha (Phage_pRha)
38024	37638	Reverse	Bacterial regulatory proteins, luxR family
39898	39107	Reverse	HTH-type transcriptional regulator PrtR

Table 93: Functional genes identified for vB_Pae_BR78a.

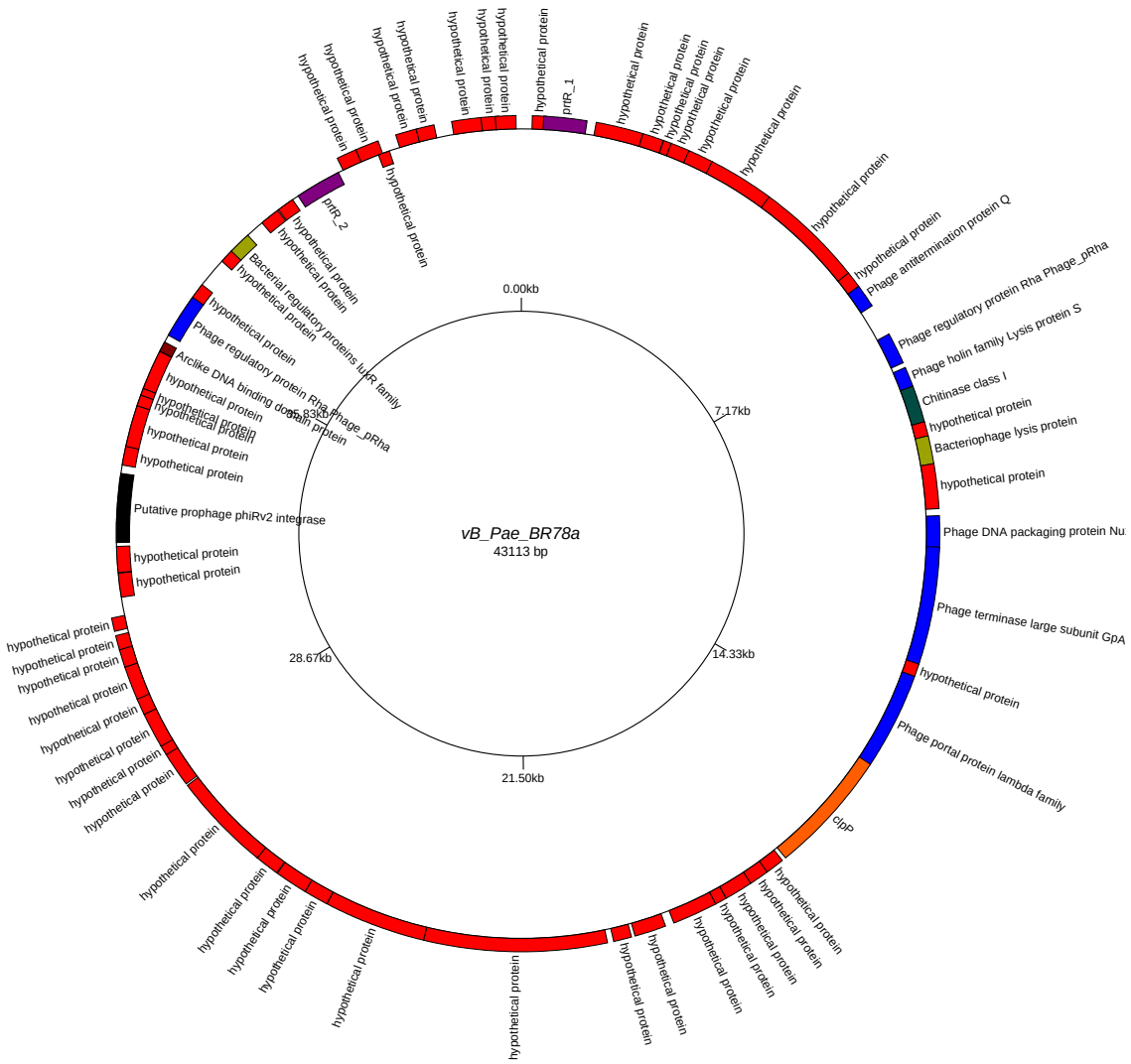


Figure 93: Genome map of vB_Pae_BR78a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR79a genome size was 61.7 kb, the phage was assembled using SPAdes and has a mean coverage of 494. The genome map of this phage is shown in Figure 94. The total number of genes identified was 64 of which 14 were identified with putative functions these can be seen in Table 94. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66 (GenBank: KC262634.1). The phage vB_Pae_BR79a had a 63.7 % G-C content. No tRNA gene was identified however; a CPS-53 integrase like gene was identified.

Start	End	Direction	Putative functional protein
16458	16997	Forward	Phage lysozyme
20063	21298	Forward	Putative prophage CPS-53 integrase
21532	21299	Reverse	Response regulator inhibitor for tor operon
25869	24061	Reverse	C-5 cytosine-specific DNA methylase
28496	25866	Reverse	DNA methylase
28850	28641	Reverse	LexA repressor
30067	28847	Reverse	Recombination-associated protein RdgC
30591	30094	Reverse	Single-stranded DNA-binding protein
36152	35781	Reverse	Carbon storage regulator homolog
38368	37571	Reverse	HTH-type transcriptional regulator PrtR
38476	38676	Reverse	Cro
41082	41696	Reverse	Bacteriophage Lambda NinG protein
42880	44163	Forward	Phage terminase large subunit
49679	50551	Forward	RyR domain protein

Table 94: Functional genes identified for vB_Pae_BR79a.

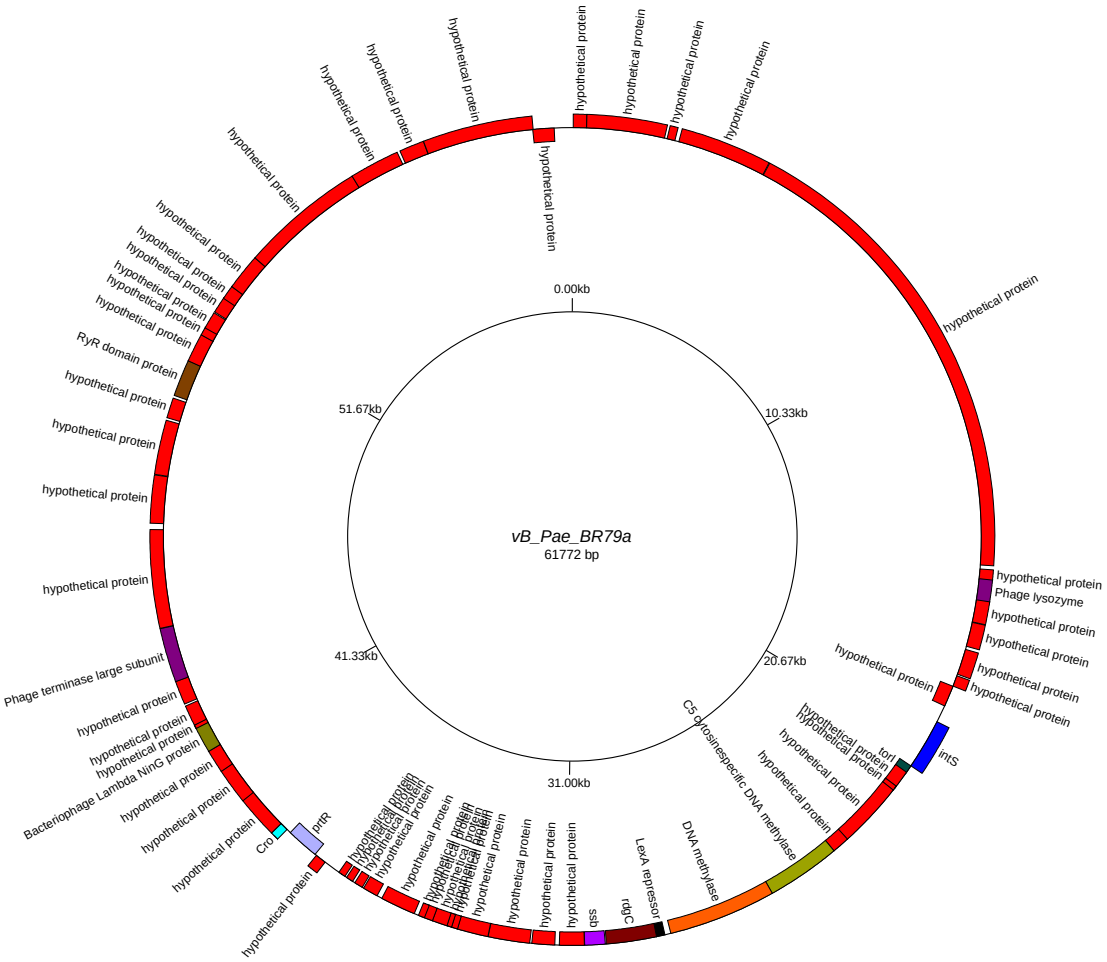


Figure 94: Genome map of vB_Pae_BR79a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR80a genome size was 37.5 kb, the phage was assembled using SPAdes and had a mean coverage of 26. The genome map of this phage is shown in Figure 95. The total number of genes identified was 56 of which 8 were identified with putative functions these can be seen in Table 95. Blastn on viruses only database showed a 27.1 kb / 27.8 kb match with *Pseudomonas* phage JBD24 (GenBank: NC_020203.1). The phage vB_Pae_BR80a had a 63.9 % G-C content. No tRNA gene was identified however; several genes associated with structural proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
1066	407	Reverse	putative HTH-type transcriptional regulator
1307	1663	Forward	DNA-binding transcriptional regulator Nlp
2826	4895	Forward	Mu DNA-binding domain protein
10282	10731	Forward	Mor transcription activator family protein
17189	18475	Forward	Phage Mu protein F like protein
18475	18942	Forward	Phage virion morphogenesis family protein
19980	21077	Forward	Mu-like prophage I protein
21081	21995	Forward	Mu-like prophage major head subunit gpT

Table 95: Functional genes identified for vB_Pae_BR80a.

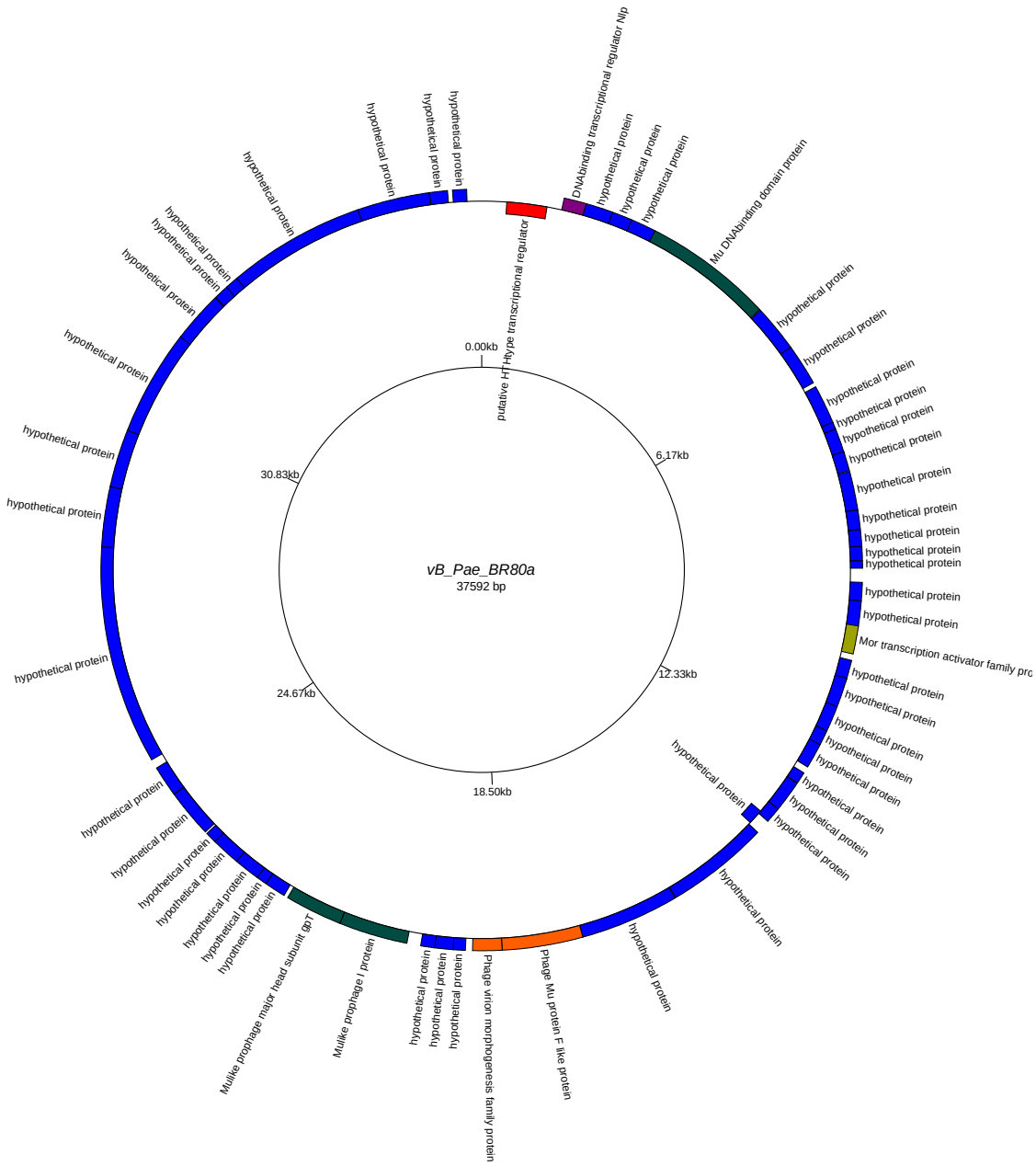


Figure 95: Genome map of vB_Pae_BR80a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The potential incomplete temperate phage vB_Pae_BR85a genome size was 29 kb, the phage was assembled using SPAdes and has a mean coverage of 22. The genome map of this phage is shown in Figure 96. The total number of genes identified was 46 of which 7 were identified with putative functions these can be seen in Table 96. Blastn on viruses only database showed a 2.5 kb / 2.6 kb match with *Pseudomonas* phage B3 (GenBank: NC_006548.1). The phage vB_Pae_BR85a had a 63.3 % G-C content. No tRNA gene was identified however; several genes associated with regulatory proteins were identified and an Integrase gene was identified.

Start	End	Direction	Putative functional protein
477	118	Reverse	Mor transcription activator family protein
7996	6212	Reverse	Integrase core domain protein
11005	11394	Forward	Helix-turn-helix domain protein
13266	13895	Forward	Soluble lytic murein transglycosylase precursor
15961	17664	Forward	Terminase-like family protein
19142	20392	Forward	Phage Mu protein F like protein
20389	20961	Forward	Phage virion morphogenesis family protein

Table 96: Functional genes identified for vB_Pae_BR85a.

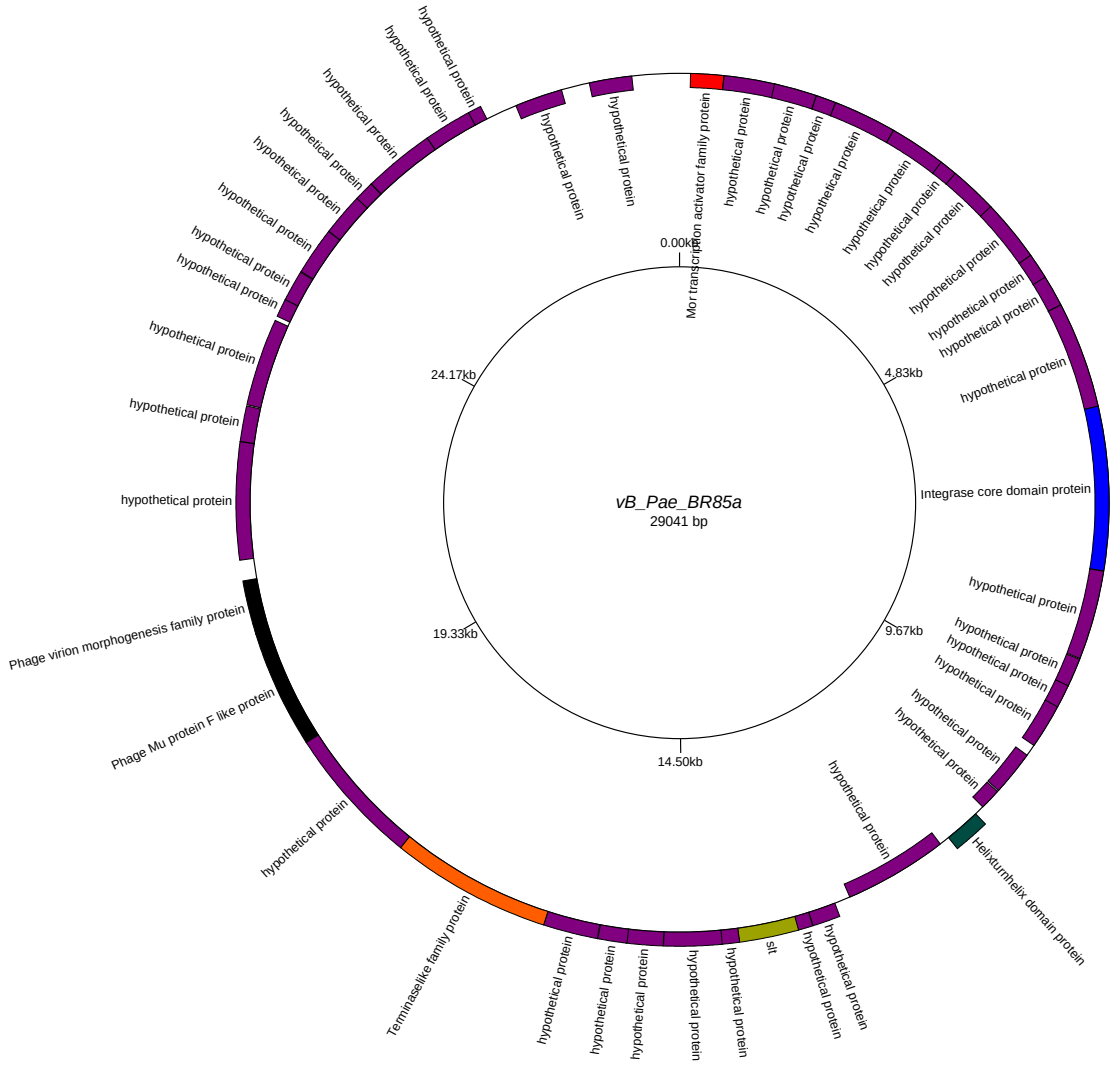


Figure 96: Genome map of vB_Pae_BR85a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR85b genome size was 38.4 kb, the phage was assembled using SPAdes and has a mean coverage of 75. The genome map of this phage is shown in Figure 97. The total number of genes identified was 59 of which 11 were identified with putative functions these can be seen in Table 97. Blastn on viruses only database showed a 17.4 kb / 18.2 kb match with *Pseudomonas* phage D3112 (GenBank: NC_005178.1). The phage vB_Pae_BR85b had a 64.4 % G-C content. No tRNA gene was identified however; Mu-like phage structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
1124	411	Reverse	HTH-type transcriptional regulator PrtR
1294	1650	Forward	DNA-binding transcriptional regulator Nlp
2352	4346	Forward	Mu DNA-binding domain protein
6559	7077	Forward	Bacteriophage Mu Gam like protein
9881	10330	Forward	Mor transcription activator family protein
16799	18085	Forward	Phage Mu protein F like protein
18085	18552	Forward	Phage virion morphogenesis family protein
19571	20677	Forward	Mu-like prophage I protein
21113	22021	Forward	Mu-like prophage major head subunit gpT
25053	28613	Forward	Prophage tail length tape measure protein
37470	38102	Forward	Succinate-semialdehyde dehydrogenase [NADP(+)]

Table 97: Functional genes identified for vB_Pae_BR85b.

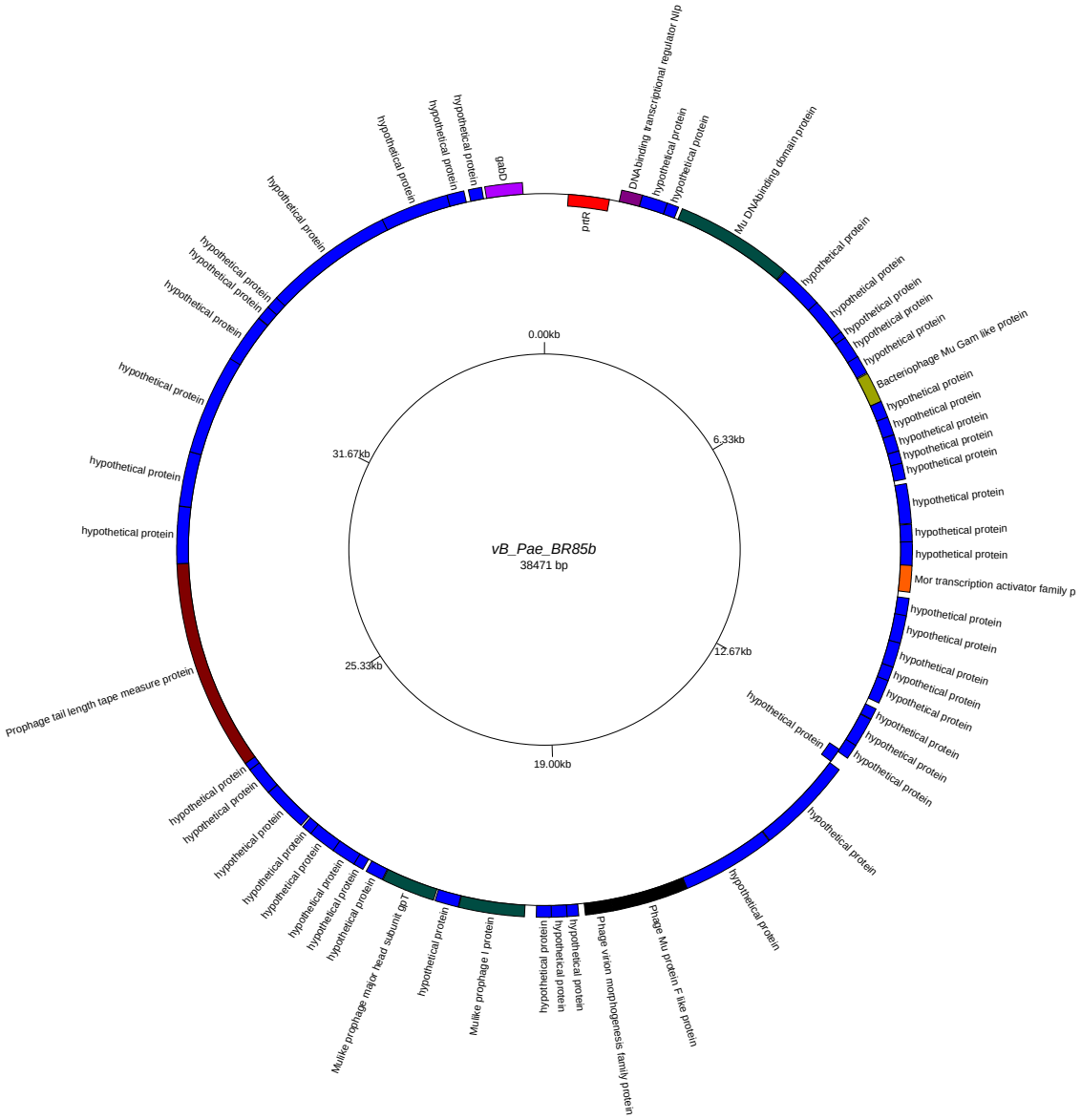


Figure 97: Genome map of vB_Pae_BR85b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR85c genome size was 52.7 kb, the phage was assembled using SPAdes and extended using PriceTI. The genome map of this phage is shown in Figure 98. The total number of genes identified was 90 of which 21 were identified with putative functions these can be seen in Table 98. Blastn on viruses only database showed a 4.9 kb / 5.1 kb match with *Pseudomonas* phage vB_PaeS_PMG1 (GenBank: NC_016765.1). The phage vB_Pae_BR85c had a 58.2 % G-C content. No tRNA gene was identified however; genes associated with structural, lysogenic and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
404	1789	Forward	Replicative DNA helicase
4203	4850	Forward	Bacteriophage Lambda NinG protein
8932	9243	Forward	HNH endonuclease
9841	11532	Forward	Phage Terminase
11686	12948	Forward	Phage portal protein
13080	13970	Forward	ATP-dependent Clp protease proteolytic subunit
13967	15154	Forward	Phage capsid family protein
16219	16575	Forward	Phage head-tail joining protein
16780	17502	Forward	P63C domain protein
21094	23598	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
31515	33161	Forward	D-glucuronyl C5-epimerase C-terminus
34555	33656	Reverse	BRO family, N-terminal domain
35158	35592	Forward	Lysozyme RrrD
37668	36712	Reverse	Tyrosine recombinase XerC
38021	37707	Reverse	Helix-turn-helix domain protein
44101	43766	Reverse	LytTr DNA-binding domain protein
44724	44098	Reverse	YqaJ-like viral recombinase domain protein
45480	44728	Reverse	ERF superfamily protein
51085	50420	Reverse	putative HTH-type transcriptional regulator
51375	51878	Forward	Bacteriophage CII protein
51882	52478	Forward	T5orf172 domain protein

Table 98: Functional genes identified for vB_Pae_BR85c.

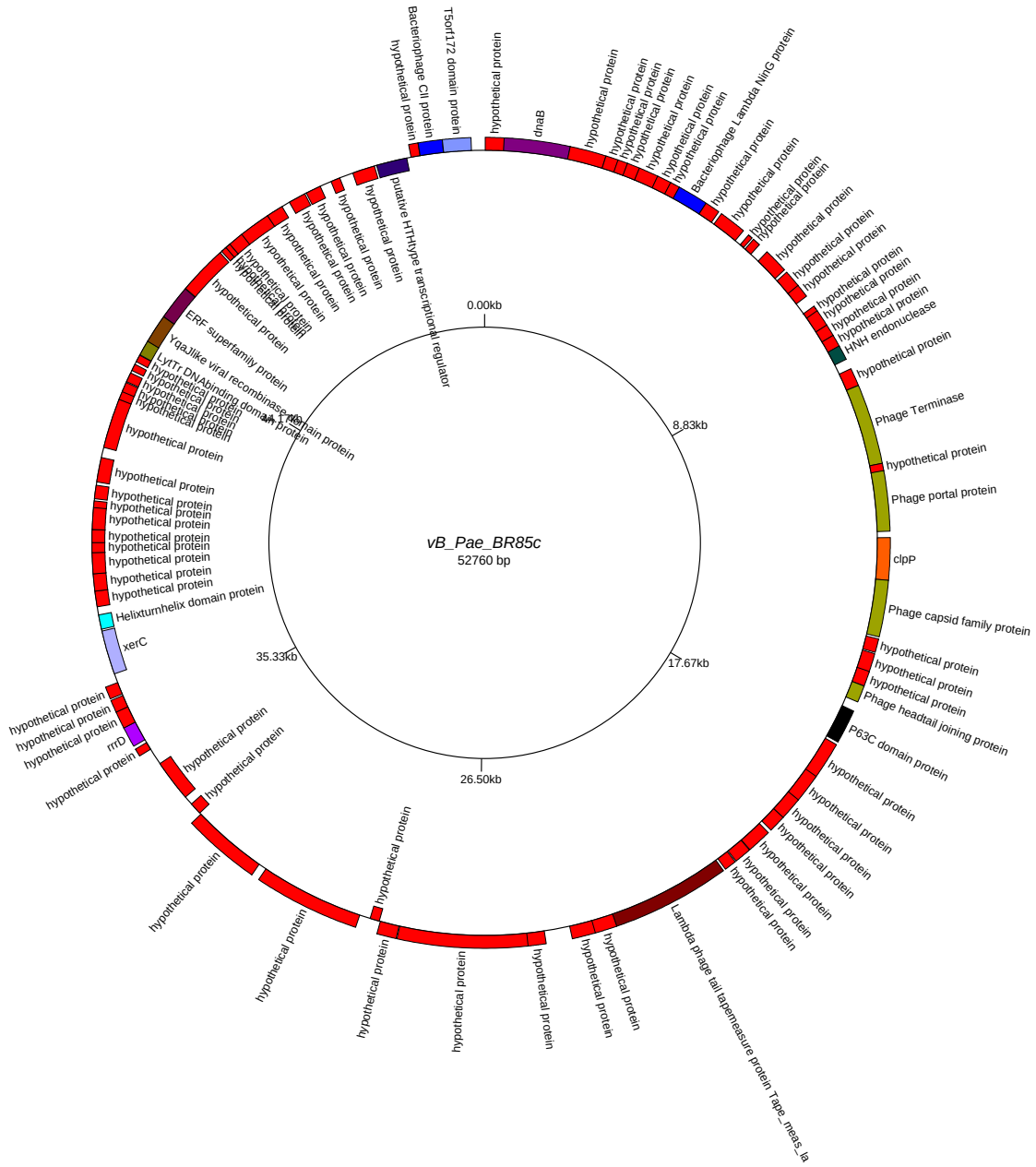


Figure 98: Genome map of vB_Pae_BR85c produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR86a genome size was 39.8 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 99. The total number of genes identified was 62 of which 13 were identified with putative functions these can be seen in Table 99. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR86a had a 61.4 % G-C content. No tRNA gene was identified however; a phage integrase like, structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
5996	4944	Reverse	Phage integrase family protein
9172	8660	Reverse	Phage regulatory protein Rha (Phage_pRha)
10577	10191	Reverse	Bacterial regulatory proteins, luxR family
12451	11660	Reverse	HTH-type transcriptional regulator PrtR
20174	20563	Forward	Phage antitermination protein Q
21102	21629	Forward	Phage regulatory protein Rha (Phage_pRha)
21714	22046	Forward	Phage holin family (Lysis protein S)
22043	22660	Forward	Chitinase class I
22896	23366	Forward	Bacteriophage lysis protein
24226	24771	Forward	Phage DNA packaging protein Nu1
24743	26707	Forward	Phage terminase large subunit (GpA)
26913	28559	Forward	Phage portal protein, lambda family
28531	30612	Forward	ATP-dependent Clp protease proteolytic subunit

Table 99: Functional genes identified for vB_Pae_BR86a.

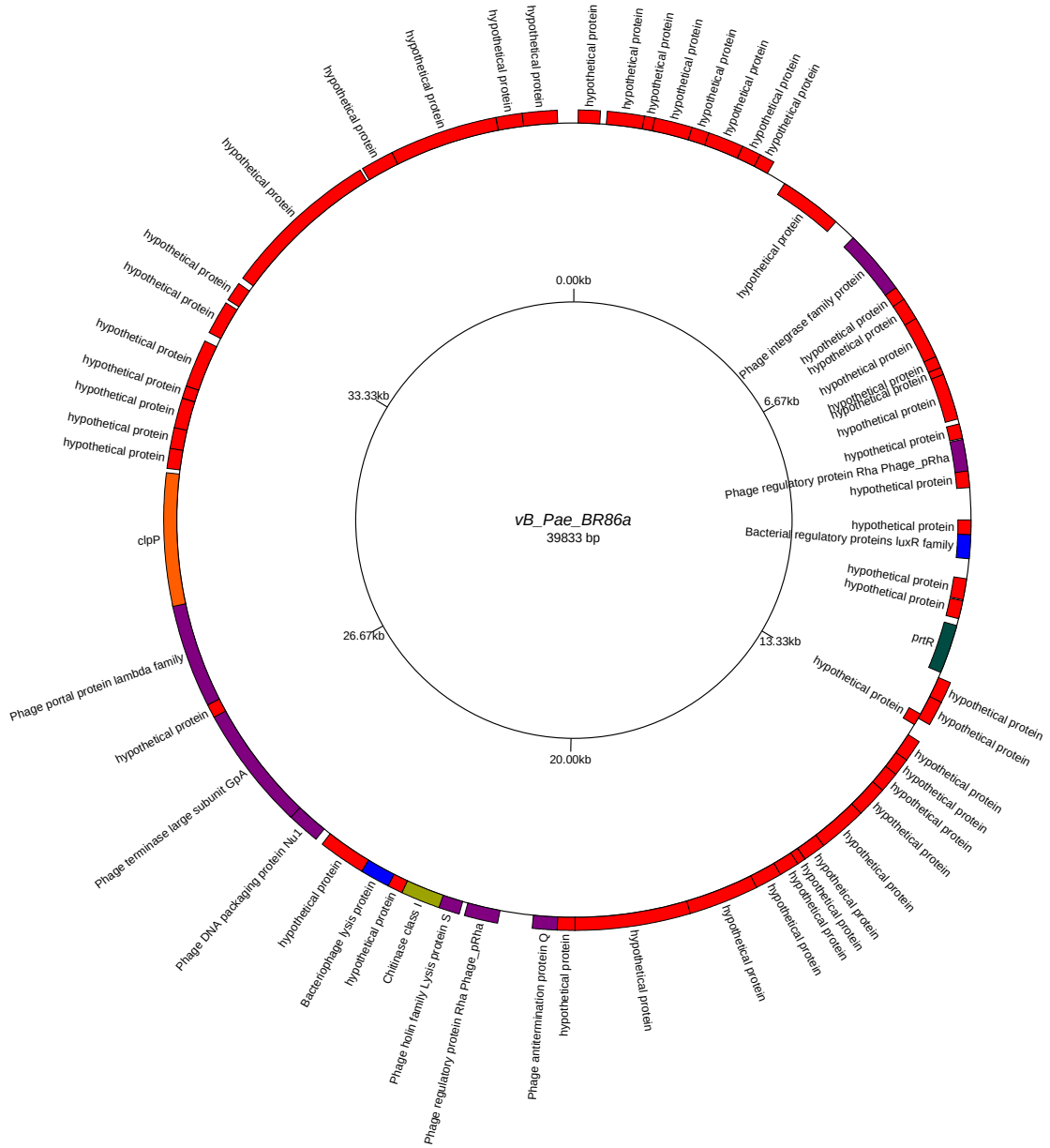


Figure 99: Genome map of vB_Pae_BR86a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR88a genome size was 39.8 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 100. The total number of genes identified was 58 of which 26 were identified with putative functions these can be seen in Table 100. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR88a had a 61.4 % G-C content. No tRNA gene was identified however; a phage integrase like, structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
666	337	Reverse	HNH endonuclease
1373	861	Reverse	Bacteriophage lysis protein
1990	1373	Reverse	Chitinase class I
2319	1987	Reverse	Phage holin family (Lysis protein S)
2934	2404	Reverse	Phage regulatory protein Rha (Phage_pRha)
4098	3709	Reverse	Phage antitermination protein Q
5767	4370	Reverse	Replicative DNA helicase
6579	5764	Reverse	DNA replication protein DnaC
10840	10535	Reverse	Helix-turn-helix
10950	11609	Forward	LexA repressor
11758	12144	Forward	Bacterial regulatory proteins, luxR family
12456	13241	Forward	BRO family, N-terminal domain
13316	13546	Forward	Arc-like DNA binding domain protein
18024	19007	Forward	Transposase DDE domain protein
23277	19612	Reverse	Carbohydrate binding domain protein
23906	23334	Reverse	Bacteriophage lambda tail assembly protein I
25415	24657	Reverse	NlpC/P60 family protein
26164	25418	Reverse	Phage minor tail protein L
26499	26161	Reverse	Phage minor tail protein
29774	26499	Reverse	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
32472	32146	Reverse	Phage head-tail joining protein

32795	32472	Reverse	Phage gp6-like head-tail connector protein
34264	33050	Reverse	Phage capsid family protein
34905	34261	Reverse	Caudovirus prohead protease
36112	34889	Reverse	Phage portal protein
37794	36115	Reverse	Phage Terminase

Table 100: Functional genes identified for vB_Pae_BR88a.

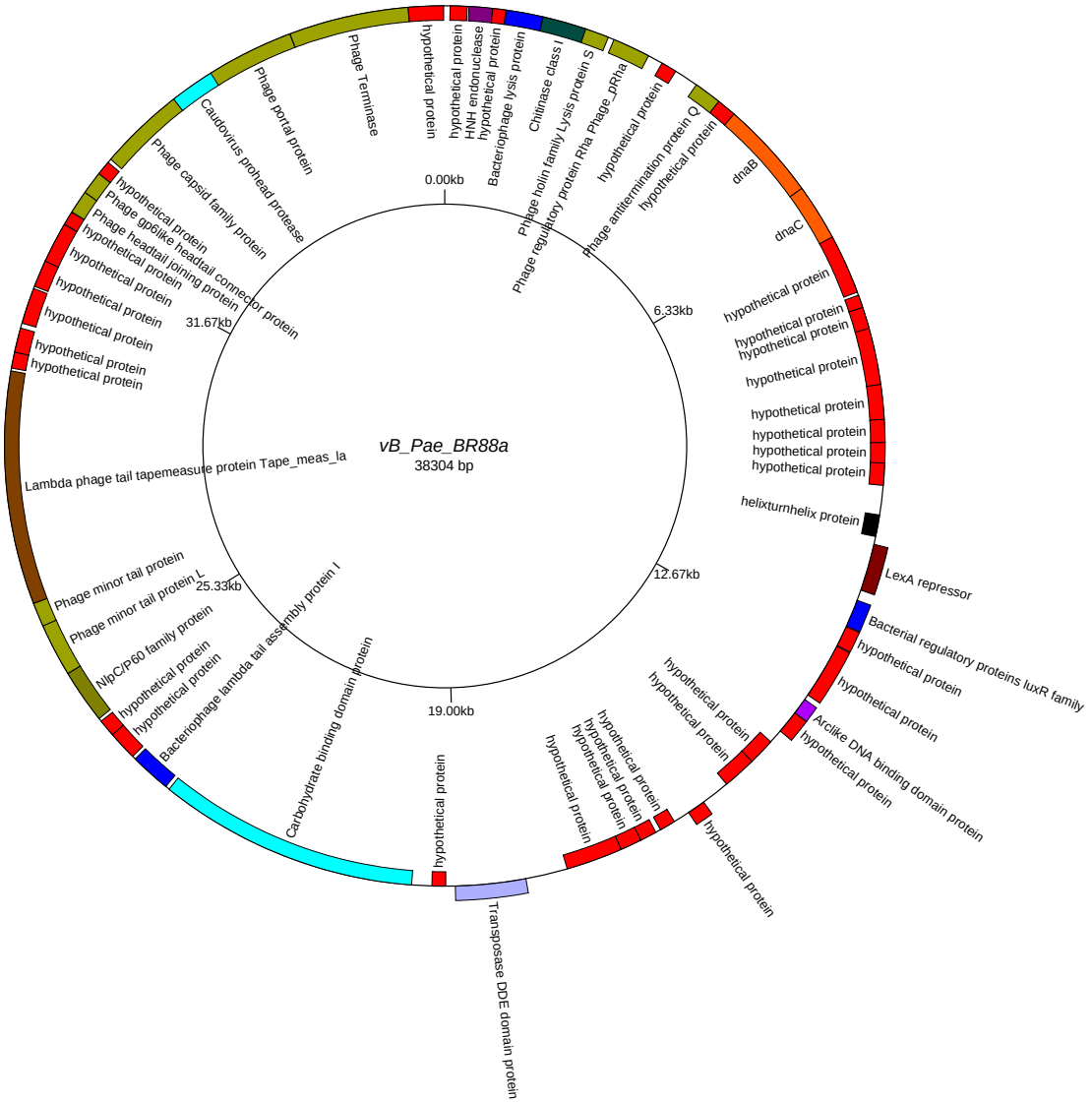


Figure 100: Genome map of vB_Pae_BR88a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR88b genome size was 45.4 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 101. The total number of genes identified was 69 of which 15 were identified with putative functions these can be seen in Table 101. Blastn on viruses only database showed an 8.8 kb / 9 kb match with *Pseudomonas* phage phi297 (GenBank: NC_016762.1). The phage vB_Pae_BR88b had a 58.9 % G-C content. No tRNA gene was identified however; CII, Ig-like domains and recombinase genes were identified.

Start	End	Direction	Putative functional protein
2122	3240	Forward	Phage terminase large subunit
4773	6503	Forward	Phage Mu protein F like protein
11388	12383	Forward	Bacterial Ig-like domain (group 2)
16469	17029	Forward	AP2 domain protein
23924	25570	Forward	D-glucuronyl C5-epimerase C-terminus
26910	26065	Reverse	BRO family, N-terminal domain
27568	28002	Forward	Lysozyme RrrD
30078	29122	Reverse	Tyrosine recombinase XerC
30431	30117	Reverse	Helix-turn-helix domain protein
36511	36176	Reverse	LytTr DNA-binding domain protein
37134	36508	Reverse	YqaJ-like viral recombinase domain protein
37890	37138	Reverse	ERF superfamily protein
43495	42830	Reverse	putative HTH-type transcriptional regulator
43785	44288	Forward	Bacteriophage CII protein
44292	44888	Forward	T5orf172 domain protein

Table 101: Functional genes identified for vB_Pae_BR88b.

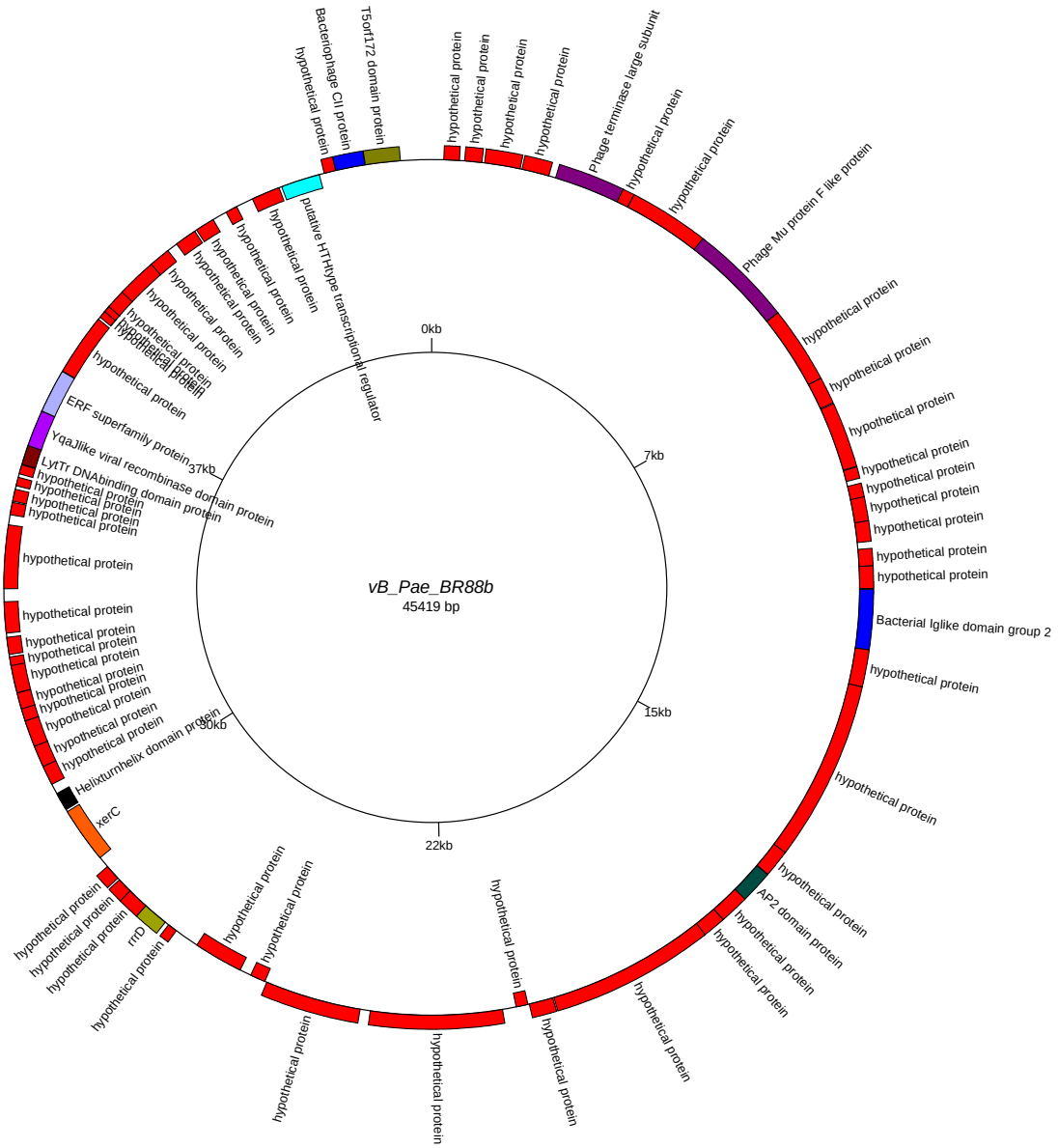


Figure 101: Genome map of vB_Pae_BR88b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR88c genome size was 39.7 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 102. The total number of genes identified was 57 of which 9 were identified with putative functions these can be seen in Table 102. Blastn on viruses only database showed a 7.1 kb / 7.6 kb match with *Pseudomonas* phage B3 (GenBank: NC_006548.1). The phage vB_Pae_BR88c had a 63.5 % G-C content. No tRNA gene was identified however; several genes associated with regulatory proteins and an Integrase protein was identified.

Start	End	Direction	Putative functional protein
10263	8932	Reverse	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
12698	10362	Reverse	Prophage tail length tape measure protein
19287	18715	Reverse	Phage virion morphogenesis family protein
20534	19284	Reverse	Phage Mu protein F like protein
23715	22012	Reverse	Terminase-like family protein
26410	25781	Reverse	Soluble lytic murein transglycosylase precursor
28671	28282	Reverse	Helix-turn-helix domain protein
31680	33464	Forward	Integrase core domain protein
39199	39558	Forward	Mor transcription activator family protein

Table 102: Functional genes identified for vB_Pae_BR88c.

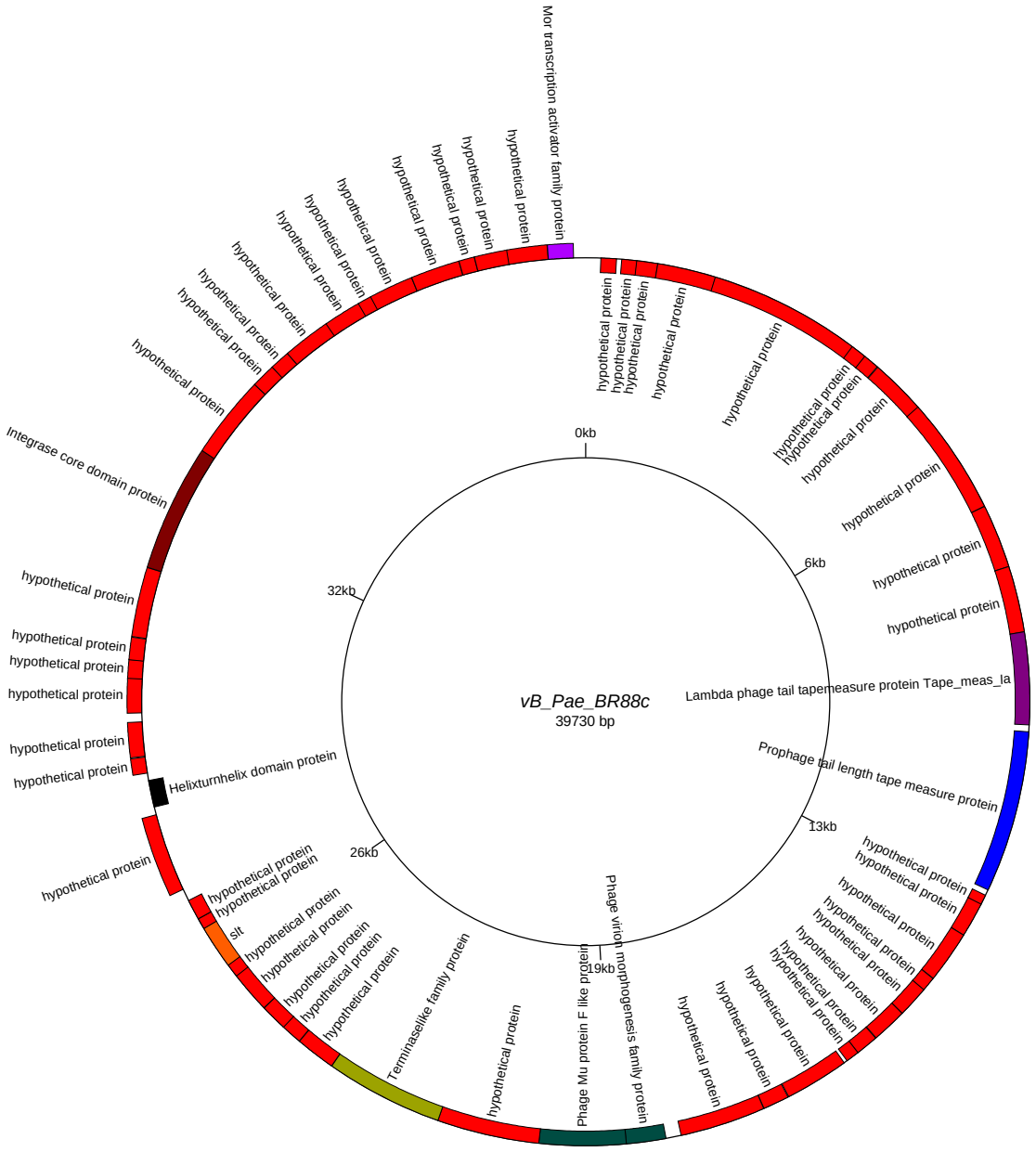


Figure 102: Genome map of vB_Pae_BR88c produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The potentially incomplete temperate phage vB_Pae_BR88d genome size was 27.5 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 103. The total number of genes identified was 49 of which 10 were identified with putative functions these can be seen in Table 103. Blastn on viruses only database showed a 17.1 kb / 27.8 kb match with *Pseudomonas* phage JBD24 (GenBank: NC_020203.1). The phage vB_Pae_BR88d had a 64.4 % G-C content. No tRNA gene was identified however; several genes associated with structural proteins were identified.

Start	End	Direction	Putative functional protein
2435	1002	Reverse	Prophage tail length tape measure protein
6375	5467	Reverse	Mu-like prophage major head subunit gpT
7917	6811	Reverse	Mu-like prophage I protein
9403	8936	Reverse	Phage virion morphogenesis family protein
10689	9403	Reverse	Phage Mu protein F like protein
17607	17158	Reverse	Mor transcription activator family protein
20929	20411	Reverse	Bacteriophage Mu Gam like protein
25136	23142	Reverse	Mu DNA-binding domain protein
26194	25838	Reverse	DNA-binding transcriptional regulator Nlp
26364	27077	Forward	HTH-type transcriptional regulator PrtR

Table 103: Functional genes identified for vB_Pae_BR88d.

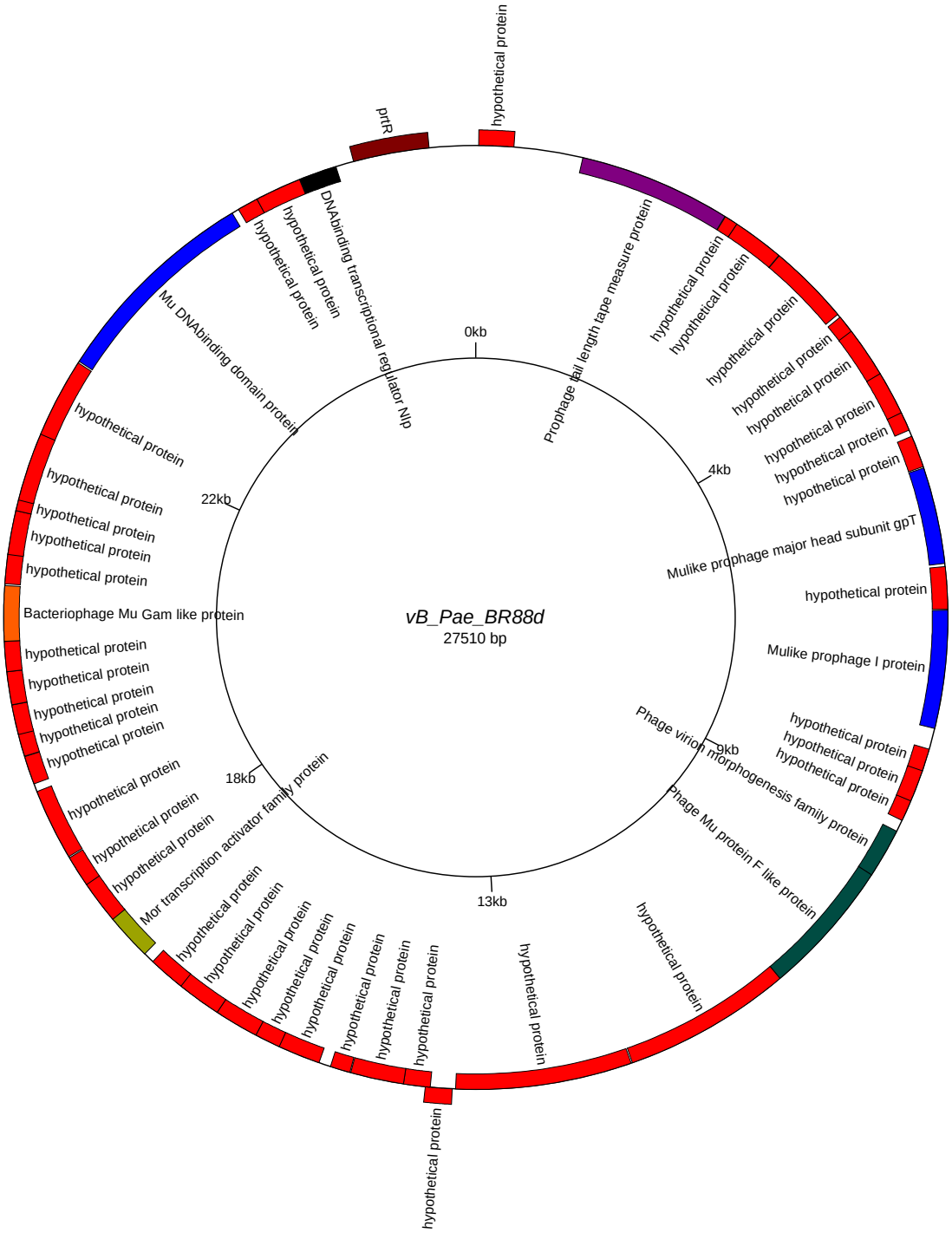


Figure 103: Genome map of vB_Pae_BR88d produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR89a genome size was 41.8 kb, the phage was assembled using SPAdes and was extended using PriceTI. The genome map of this phage is shown in Figure 104. The total number of genes identified was 66 of which 12 were identified with putative functions these can be seen in Table 104. Blastn on viruses only database showed a 9.7 kb / 10.1 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR89a had a 61.6 % G-C content. No tRNA gene was identified however; an integrase, structural and regulatory genes were identified.

Start	End	Direction	Putative functional protein
4350	5528	Forward	Putative prophage phiRv2 integrase
22932	20851	Reverse	ATP-dependent Clp protease proteolytic subunit
24550	22904	Reverse	Phage portal protein, lambda family
26720	24756	Reverse	Phage terminase large subunit (GpA)
27237	26692	Reverse	Phage DNA packaging protein Nu1
28564	28094	Reverse	Bacteriophage lysis protein
29417	28800	Reverse	Chitinase class I
29746	29414	Reverse	Phage holin family (Lysis protein S)
30358	29831	Reverse	Phage regulatory protein Rha (Phage_pRha)
31286	30897	Reverse	Phage antitermination protein Q
39008	39799	Forward	HTH-type transcriptional regulator PrtR
40882	41133	Forward	Bacterial regulatory proteins, luxR family

Table 104: Functional genes identified for vB_Pae_BR89a.

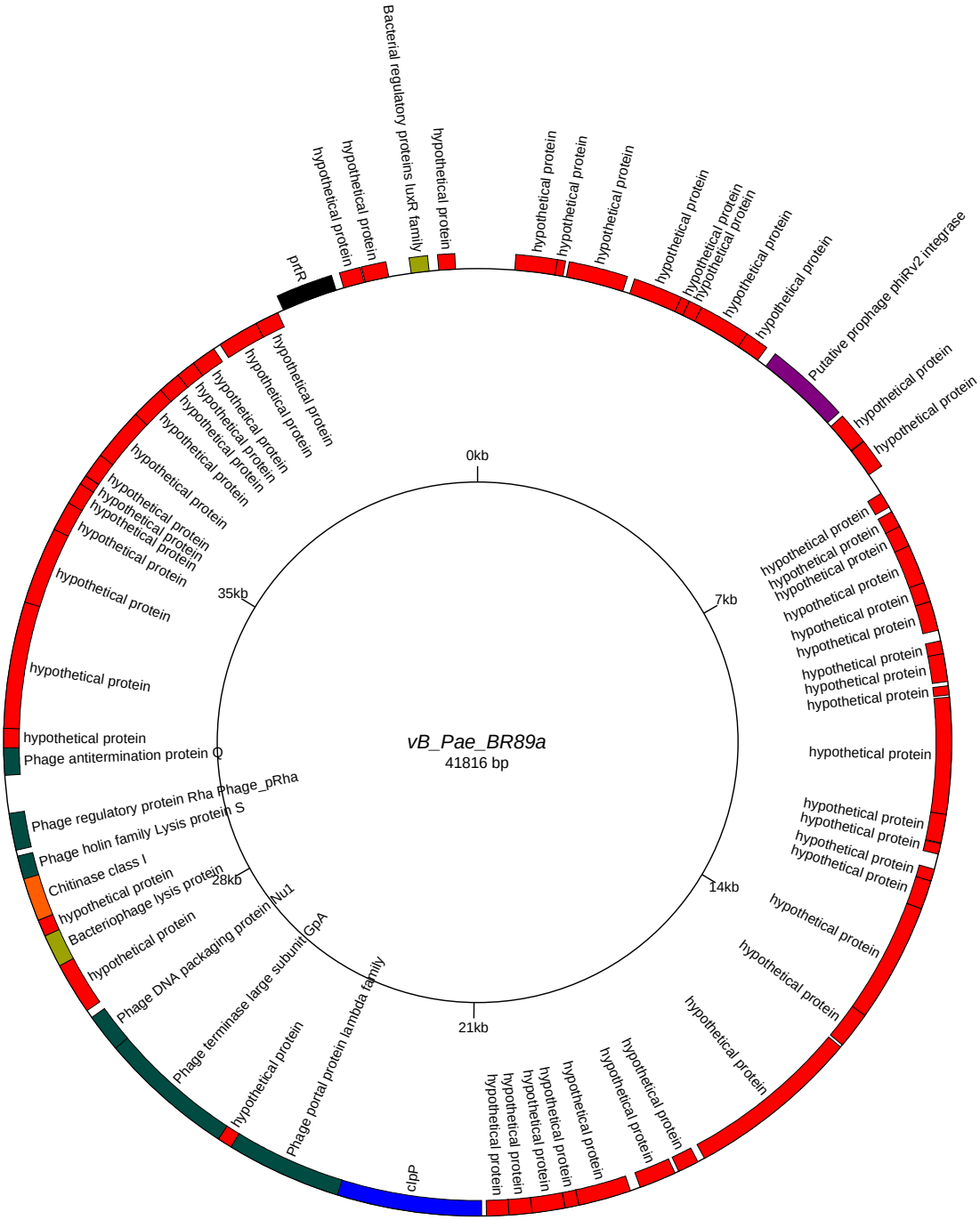


Figure 104: Genome map of vB_Pae_BR89a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR91a genome size was 61.7 kb, the phage was assembled using SPAdes and has a mean coverage of 469. The genome map of this phage is shown in Figure 105. The total number of genes identified was 62 of which 14 were identified with putative functions these can be seen in Table 105. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66 (GenBank: KC262634.1). The phage vB_Pae_BR91a had a 63.7 % G-C content. No tRNA gene was identified however; a CPS-53 integrase like gene was identified.

Start	End	Direction	Putative functional protein
4533	5030	Forward	Single-stranded DNA-binding protein
5057	6277	Forward	Recombination-associated protein RdgC
6274	6483	Forward	LexA repressor
6628	9258	Forward	DNA methylase
9255	11063	Forward	C-5 cytosine-specific DNA methylase
13592	13825	Forward	Response regulator inhibitor for tor operon
15061	13826	Reverse	Putative prophage CPS-53 integrase
18666	18127	Reverse	Phage lysozyme
47090	46218	Reverse	RyR domain protein
53889	52606	Reverse	Phage terminase large subunit
55687	55073	Reverse	Bacteriophage Lambda NinG protein
58293	58093	Reverse	Cro
58401	59198	Forward	HTH-type transcriptional regulator PrtR
60617	60988	Forward	Carbon storage regulator homolog

Table 105: Functional genes identified for vB_Pae_BR91a.

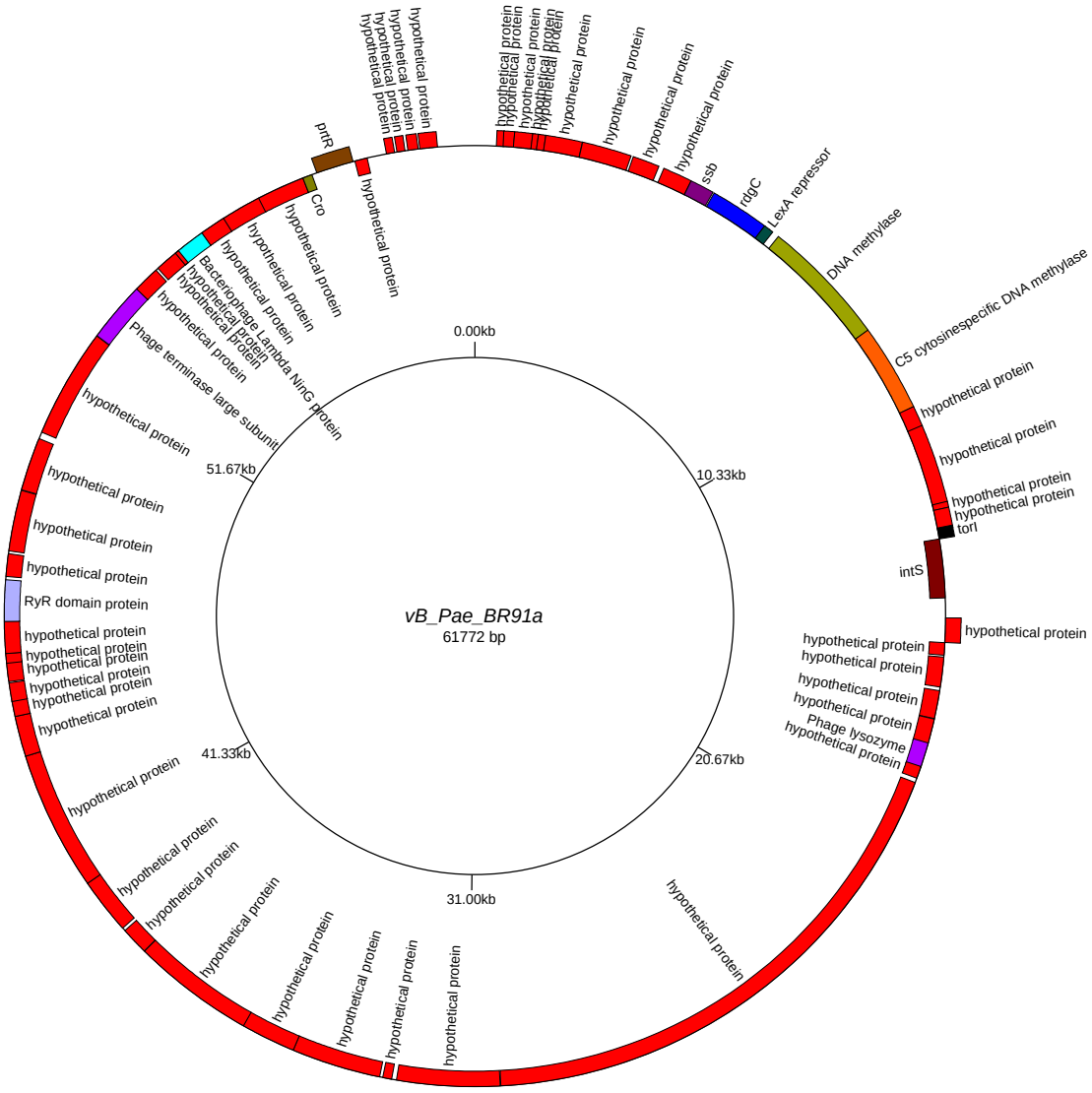


Figure 105: Genome map of vB_Pae_BR91a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR92a genome size was 61.7 kb, the phage was assembled using SPAdes and has a mean coverage of 621. The genome map of this phage is shown in Figure 106. The total number of genes identified was 62 of which 14 were identified with putative functions these can be seen in Table 106. Blastn on viruses only database showed an 11.7 kb / 12.1 kb match with *Pseudomonas* phage H66 (GenBank: KC262634.1). The phage vB_Pae_BR92a had a 63.7 % G-C content. No tRNA gene was identified however; a CPS-53 integrase like gene was identified.

Start	End	Direction	Putative functional protein
4605	5102	Forward	Single-stranded DNA-binding protein
5129	6349	Forward	Recombination-associated protein RdgC
6346	6555	Forward	LexA repressor
6700	9330	Forward	DNA methylase
9327	11135	Forward	C-5 cytosine-specific DNA methylase
13664	13897	Forward	Response regulator inhibitor for tor operon
15133	13898	Reverse	Putative prophage CPS-53 integrase
18738	18199	Reverse	Phage lysozyme
47162	46290	Reverse	RyR domain protein
53961	52678	Reverse	Phage terminase large subunit
55759	55145	Reverse	Bacteriophage Lambda NinG protein
58365	58165	Reverse	Cro
58473	59270	Forward	HTH-type transcriptional regulator PrtR
60689	61060	Forward	Carbon storage regulator homolog

Table 106: Functional genes identified for vB_Pae_BR92a.

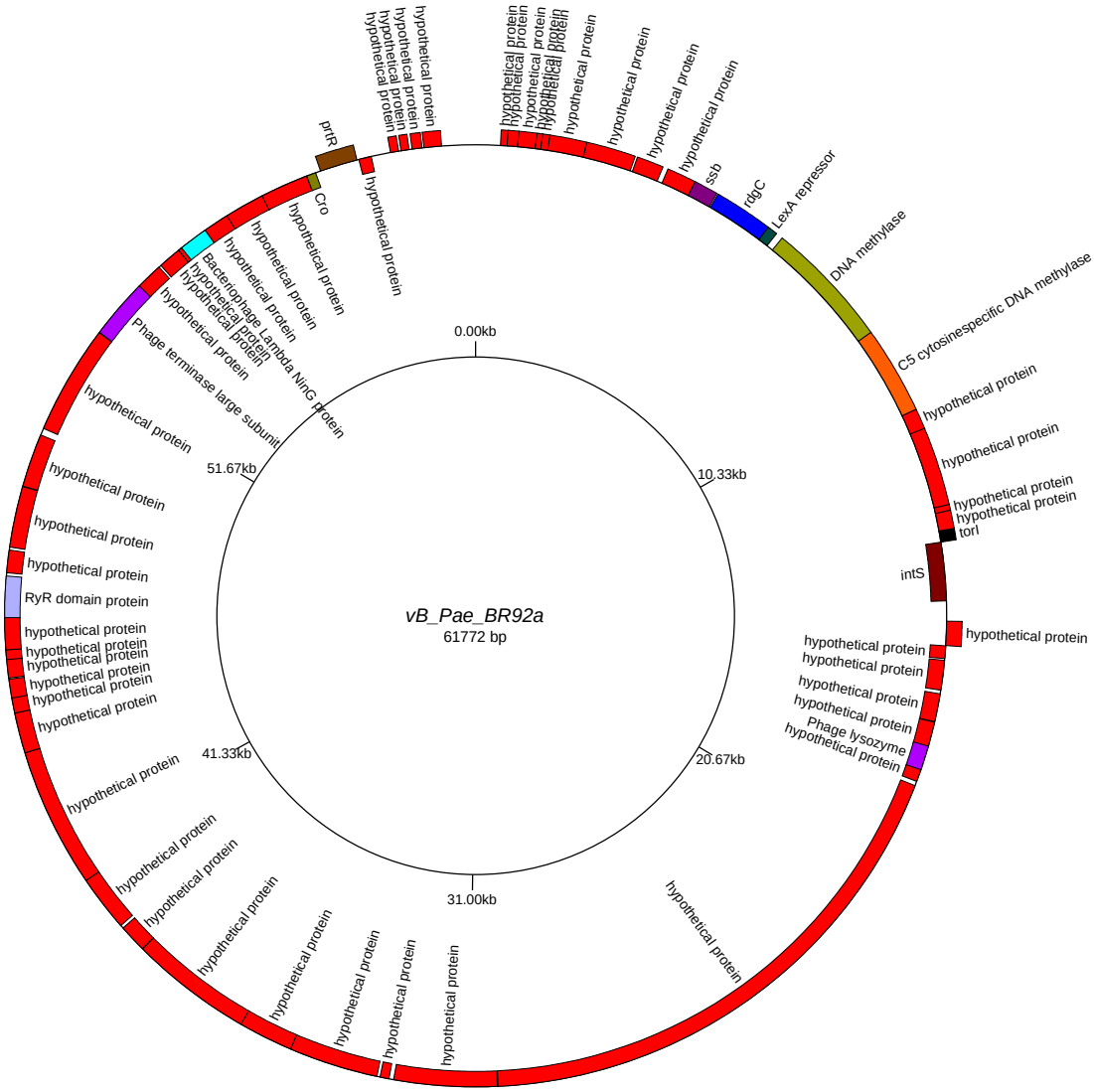


Figure 106: Genome map of vB_Pae_BR92a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Pae_BR93a genome size was 38.3 kb, the phage was assembled using SPAdes and extended using PriceTI. The genome map of this phage is shown in Figure 107. The total number of genes identified was 67 of which 14 were identified with putative functions these can be seen in Table 107. Blastn on viruses only database showed a 6.2 kb / 6.7 kb match with *Pseudomonas* phage F10 (GenBank: NC_007805.1). The phage vB_Pae_BR93a had a 60.8 % G-C content. No tRNA gene was identified however; genes associated with structural and regulatory proteins were identified. An integrase gene was also predicted.

Start	End	Direction	Putative functional protein
3360	2182	Reverse	Putative prophage phiRv2 integrase
5770	5540	Reverse	Arc-like DNA binding domain protein
6617	5850	Reverse	Phage regulatory protein Rha (Phage_pRha)
8022	7636	Reverse	Bacterial regulatory proteins, luxR family
9896	9105	Reverse	HTH-type transcriptional regulator PrtR
17619	18008	Forward	Phage antitermination protein Q
18547	19074	Forward	Phage regulatory protein Rha (Phage_pRha)
19159	19491	Forward	Phage holin family (Lysis protein S)
19488	20105	Forward	Chitinase class I
20341	20811	Forward	Bacteriophage lysis protein
31982	30930	Reverse	Phage integrase family protein
35158	34391	Reverse	Phage regulatory protein Rha (Phage_pRha)
37700	36912	Reverse	HTH-type transcriptional regulator PrtR
37803	38123	Forward	Cro

Table 107: Functional genes identified for vB_Pae_BR93a.

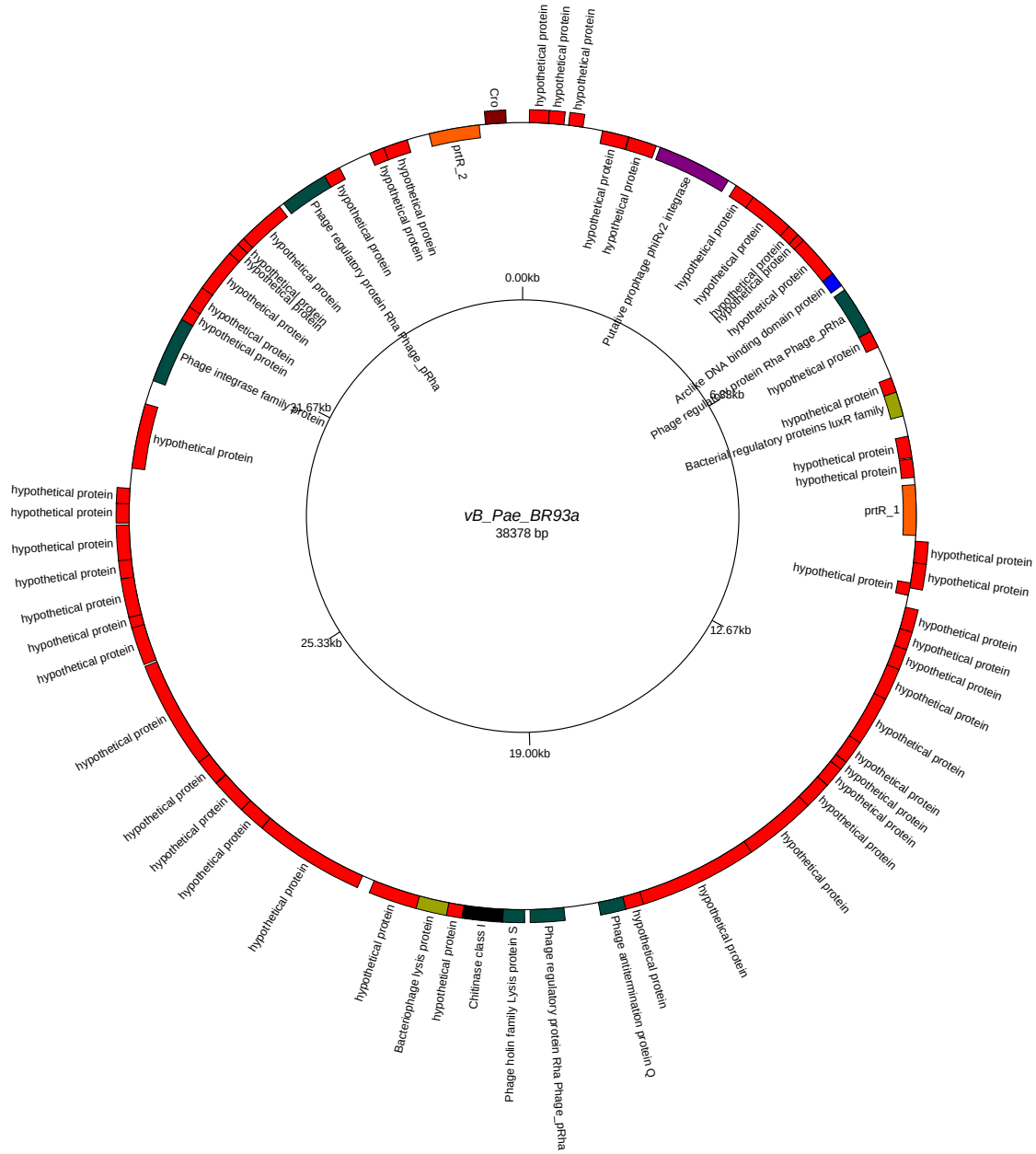


Figure 107: Genome map of vB_Pae_BR93a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

Appendix 4 *Burkholderia cepacia* complexes phage genomes

The temperate phage vB_Bcc_4a genome size was 42.9 kb, the phage was assembled using SPAdes and has a mean coverage of 7. The genome map of this phage is shown in Figure 1. The total number of genes identified was 49 of which 12 were identified with putative functions these can be seen in Table 1. Blastn on viruses only database showed a 790 bp / 1077 bp match with *Burkholderia* phage JG068 (GenBank: NC_022916.1). The phage vB_Bcc_4a had a 62.8 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
10357	8789	Reverse	Bacteriophage head to tail connecting protein
14489	12105	Reverse	DNA-dependent RNA polymerase
15530	14562	Reverse	DNA ligase
16492	16250	Reverse	MazG nucleotide pyrophosphohydrolase domain protein
18351	17974	Reverse	Recombination endonuclease VII
19276	18329	Reverse	DNA polymerase I
22766	20292	Reverse	DNA polymerase I, thermostable
24188	22950	Reverse	DnaB-like helicase C terminal domain
24997	24191	Reverse	DNA primase
31310	32347	Forward	site-specific tyrosine recombinase XerC
33410	32943	Reverse	Lysozyme RrrD
38037	36331	Reverse	Phage T7 tail fibre protein

Table 1: Functional genes identified for vB_Bcc_4a.

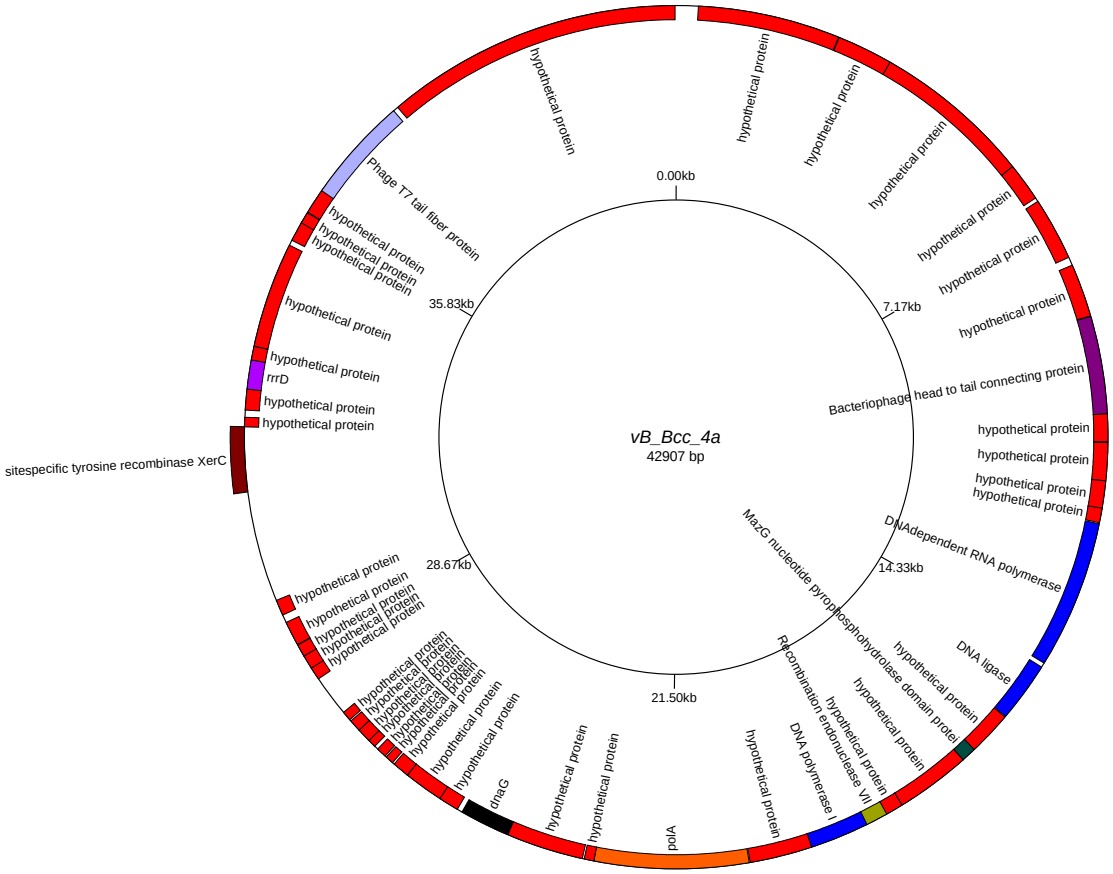


Figure 1: Genome map of vB_Bcc_4a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_4b genome size was 42.2 kb, the phage was assembled using SPAdes and has a mean coverage of 36. The genome map of this phage is shown in Figure 2. The total number of genes identified was 60 of which 24 were identified with putative functions these can be seen in Table 2. Blastn on viruses only database showed a 788 bp / 941 bp match with *Burkholderia* phage KS9 (GenBank: NC_013055.1). The phage vB_Bcc_4b had a 61.7 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
570	2558	Forward	Phage terminase large subunit (GpA)
2771	4261	Forward	Phage portal protein, lambda family
4258	5313	Forward	ATP-dependent Clp protease proteolytic subunit
7052	7579	Forward	Prophage minor tail protein Z (GPZ)
8096	8776	Forward	Phage-related baseplate assembly protein
9044	9388	Forward	Gene 25-like lysozyme
9385	10278	Forward	Baseplate J-like protein
10271	10837	Forward	Phage tail protein (Tail_P2_I)
12463	13632	Forward	Phage tail sheath protein
13643	14146	Forward	Phage tail tube protein FII
14625	17042	Forward	Phage-related minor tail protein
17052	17933	Forward	Phage P2 GpU
17908	18114	Forward	Phage Tail Protein X
18125	19177	Forward	Phage late control gene D protein (GPD)
19252	19536	Forward	Phage holin family 2
19539	20033	Forward	Phage lysozyme
20697	21488	Forward	Modification methylase DpnIIA
23137	22766	Reverse	Alkaline phosphatase synthesis transcriptional regulatory protein PhoP
26084	26983	Forward	Protease HtpX
28615	27098	Reverse	RecF/RecN/SMC N terminal domain
29023	30297	Forward	Phage integrase family protein

30529	30275	Reverse	Prophage CP4-57 regulatory protein (AlpA)
37450	39957	Forward	DNA primase TraC

Table 2: Functional genes identified for vB_Bcc_4b.

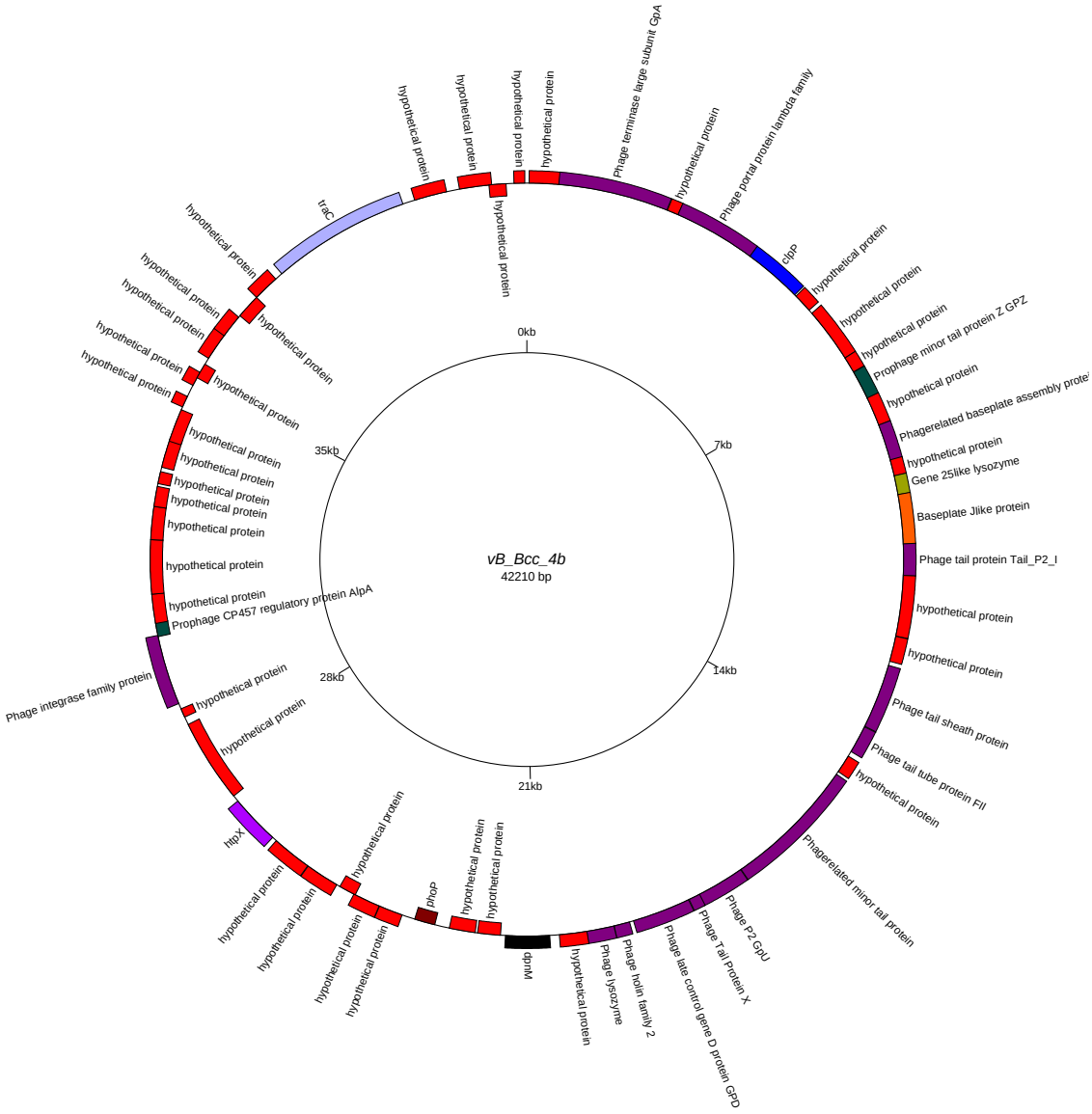
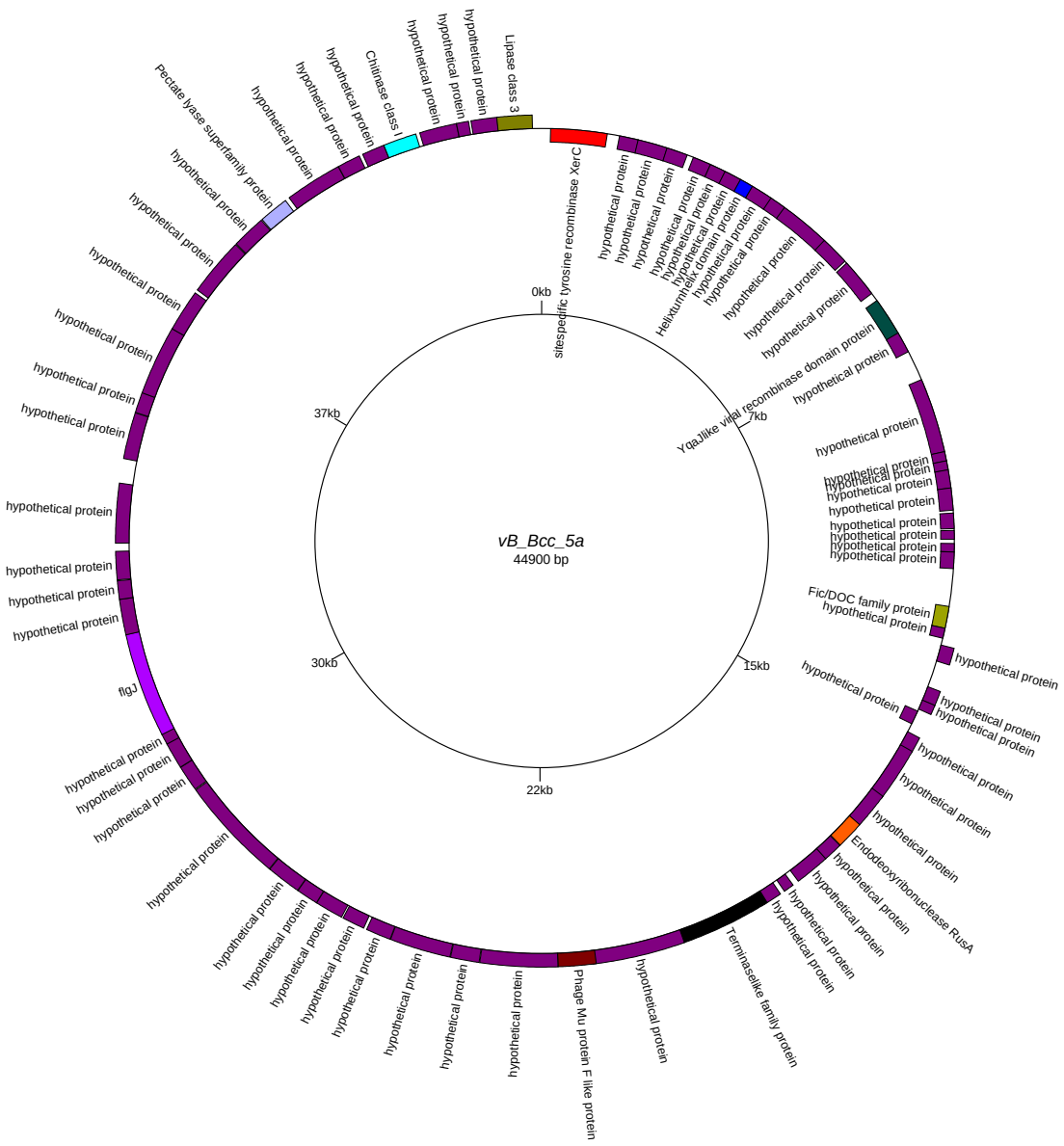


Figure 2: Genome map of vB_Bcc_4b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_5a genome size was 44.9 kb, the phage was assembled using SPAdes and has a mean coverage of 71. The genome map of this phage is shown in Figure 3. The total number of genes identified was 73 of which 11 were identified with putative functions these can be seen in Table 3. Blastn on viruses only database showed a 127 bp / 154 bp match with *Burkholderia* phage Bcep176 (GenBank: NC_007497.1). The phage vB_Bcc_5a had a 63.3 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
1142	153	Reverse	site-specific tyrosine recombinase XerC
3834	3592	Reverse	Helix-turn-helix domain protein
7456	6776	Reverse	YqaJ-like viral recombinase domain protein
12748	12365	Reverse	Fic/DOC family protein
16432	16902	Forward	Endodeoxyribonuclease RusA
18466	20013	Forward	Terminase-like family protein
21532	22173	Forward	Phage Mu protein F like protein
30304	32091	Forward	Peptidoglycan hydrolase FlgJ
39786	40292	Forward	Pectate lyase superfamily protein
42190	42756	Forward	Chitinase class I
44145	44741	Forward	Lipase (class 3)

Table 3: Functional genes identified for vB_Bcc_5a.


vB_Bcc_5a
44900 bp
0kb
7kb
15kb
22kb
30kb
37kb
Lipase class 3
hypothetical protein
hypothetical protein
hypothetical protein
Chitinase class I
hypothetical protein
hypothetical protein
hypothetical protein
Pectate lyase superfamily protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
flgJ
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Bcc_6a genome size was 45 kb, the phage was assembled using SPAdes and has a mean coverage of 12. The genome map of this phage is shown in Figure 4. The total number of genes identified was 72 of which 11 were identified with putative functions these can be seen in Table 4. Blastn on viruses only database showed a 127 bp / 154 bp match with *Burkholderia* phage Bcep176 (GenBank: NC_007497.1). The phage vB_Bcc_6a had a 63.3 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
991	1461	Forward	Endodeoxyribonuclease RusA
3025	4572	Forward	Terminase-like family protein
6091	6732	Forward	Phage Mu protein F like protein
14863	16650	Forward	Peptidoglycan hydrolase FlgJ
24351	25907	Forward	Pectate lyase superfamily protein
26754	27320	Forward	Chitinase class I
28709	29305	Forward	Lipase (class 3)
30592	29603	Reverse	site-specific tyrosine recombinase XerC
33284	33042	Reverse	Helix-turn-helix domain protein
36908	36228	Reverse	YqaJ-like viral recombinase domain protein
42200	41817	Reverse	Fic/DOC family protein

Table 4: Functional genes identified for vB_Bcc_6a.

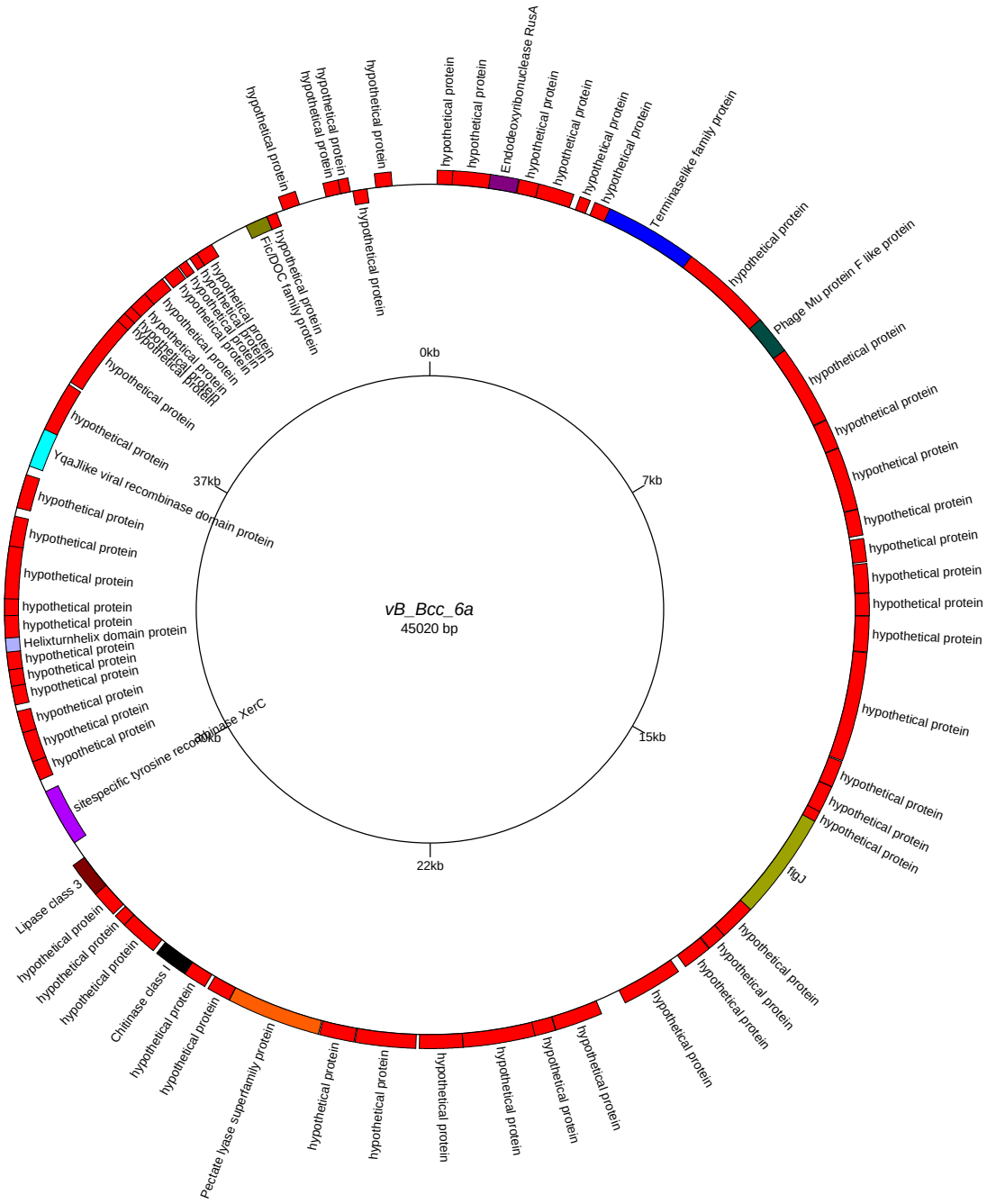


Figure 4: Genome map of vB_Bcc_6a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_6b genome size was 40.9 kb, the phage was assembled using SPAdes and has a mean coverage of 1117. The genome map of this phage is shown in Figure 5. The total number of genes identified was 54 of which 24 were identified with putative functions these can be seen in Table 5. Blastn on viruses only database showed a 36.7 kb / 36.7 kb match with *Burkholderia* phage BcepMu (GenBank: NC_005882.1). The phage vB_Bcc_6b had a 63 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
899	1735	Forward	Integrase core domain protein
2612	1800	Reverse	Small-conductance mechanosensitive channel
2892	4025	Forward	O-acetyltransferase OatA
7164	6583	Reverse	Phage tail protein (Tail_P2_I)
8308	7157	Reverse	Baseplate J-like protein
8658	8305	Reverse	Gene 25-like lysozyme
9314	8712	Reverse	Phage-related baseplate assembly protein
10516	9311	Reverse	Phage late control gene D protein (GPD)
10713	10504	Reverse	Phage Tail Protein X
11603	10713	Reverse	Phage P2 GpU
14145	11605	Reverse	Phage-related minor tail protein
15525	15001	Reverse	Phage tail tube protein FII
16961	15528	Reverse	Phage tail sheath protein
17671	17207	Reverse	Gp37 protein
21361	20834	Reverse	Phage virion morphogenesis family protein
22194	21358	Reverse	Phage Mu protein F like protein
25161	23659	Reverse	Terminase-like family protein
27576	26965	Reverse	membrane-bound lytic transglycosylase F
30255	29839	Reverse	Helix-turn-helix domain protein
31507	31809	Forward	IclR helix-turn-helix domain protein
33014	34813	Forward	Integrase core domain protein

37617	37889	Forward	DNA-binding protein HU
38745	39116	Forward	Mor transcription activator family protein
40932	39481	Reverse	Squalene--hopene cyclase

Table 5: Functional genes identified for vB_Bcc_6b.

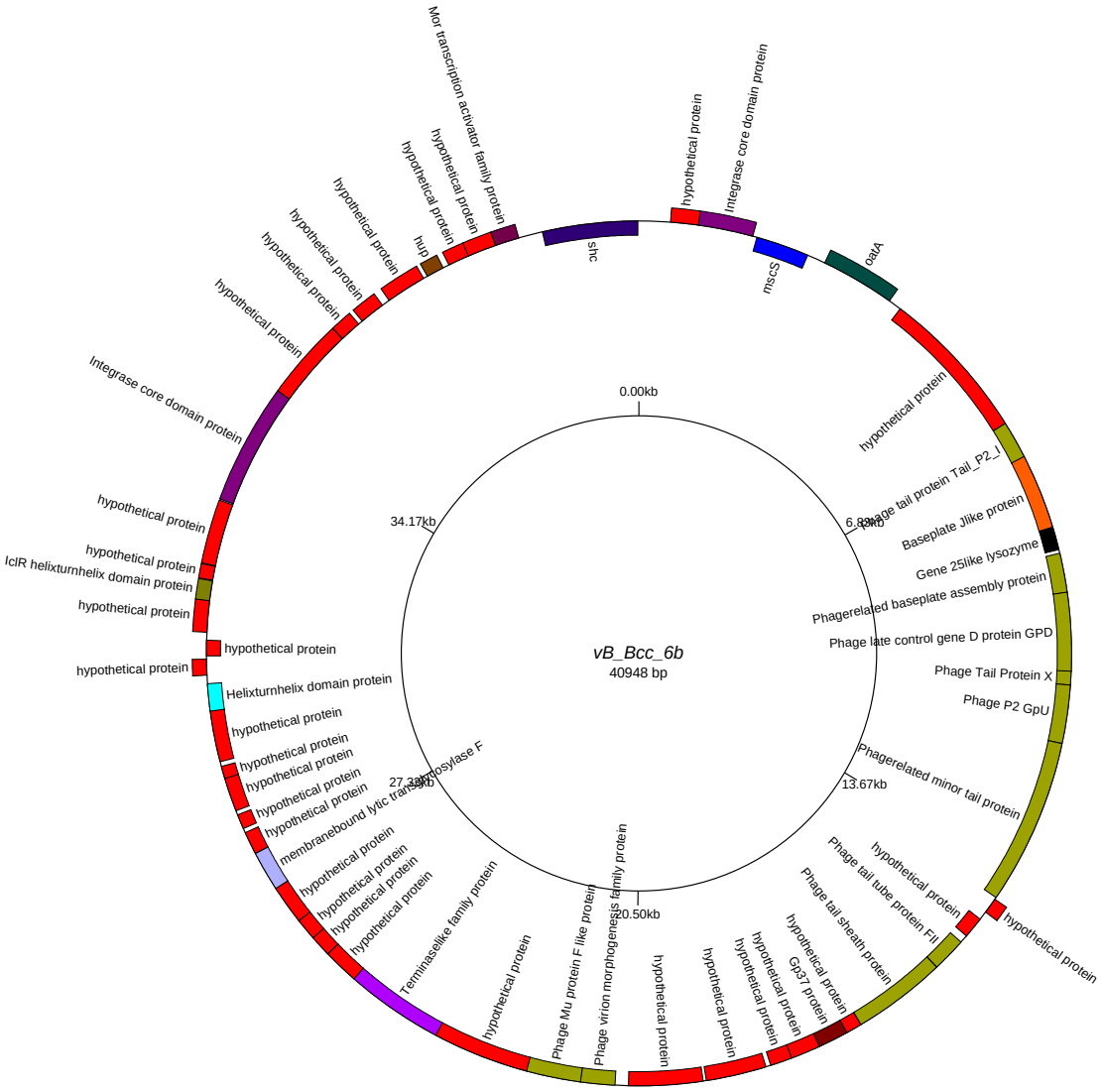


Figure 5: Genome map of vB_Bcc_6b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_6c genome size was 37.8 kb, the phage was assembled using SPAdes and has a mean coverage of 21. The genome map of this phage is shown in Figure 6. The total number of genes identified was 54 of which 14 were identified with putative functions these can be seen in Table 6. Blastn on viruses only database showed a 37.6 kb / 37.6 kb match with *Burkholderia* phage KS10 (GenBank: NC_011216.1). The phage vB_Bcc_6c had a 62.8% G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
1458	226	Reverse	Phage Mu protein F like protein
7158	6487	Reverse	membrane-bound lytic transglycosylase F
9330	8860	Reverse	Helix-turn-helix domain protein
10692	12317	Forward	Integrase core domain protein
15480	15752	Forward	DNA-binding protein HU-beta
21703	22134	Forward	Phage virion morphogenesis family protein
23042	24520	Forward	Phage tail sheath protein
25020	25574	Forward	Mu-like prophage FluMu protein gp41
25623	28058	Forward	Phage-related minor tail protein
28058	29428	Forward	DNA circulation protein N-terminus
29434	30591	Forward	Phage late control gene D protein (GPD)
30591	31112	Forward	Bacteriophage Mu Gp45 protein
31197	31778	Forward	Phage protein GP46
31775	32896	Forward	Baseplate J-like protein

Table 6: Functional genes identified for vB_Bcc_6c.

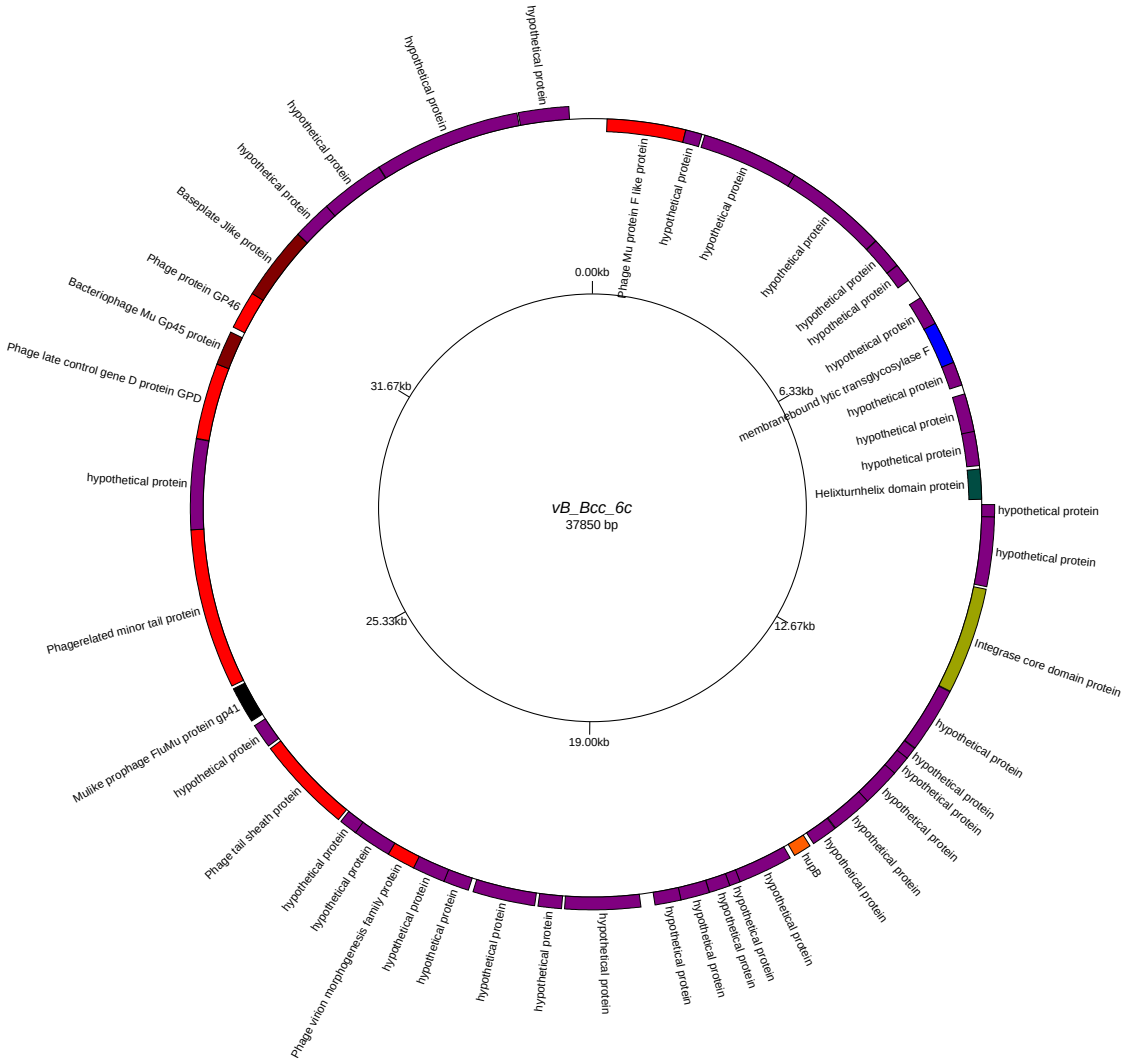


Figure 6: Genome map of vB_Bcc_6c produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_7a genome size was 37 kb, the phage was assembled using SPAdes and has a mean coverage of 1117. The genome map of this phage is shown in Figure 7. The total number of genes identified was 50 of which 21 were identified with putative functions these can be seen in Table 7. Blastn on viruses only database showed a 36.7 kb / 36.7 kb match with *Burkholderia* phage BcepMu (GenBank: NC_005882.1). The phage vB_Bcc_7a had a 62.9 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
335	1468	Forward	O-acetyltransferase OatA
4607	4026	Reverse	Phage tail protein (Tail_P2_I)
5751	4600	Reverse	Baseplate J-like protein
6101	5748	Reverse	Gene 25-like lysozyme
6757	6155	Reverse	Phage-related baseplate assembly protein
7959	6754	Reverse	Phage late control gene D protein (GPD)
8156	7947	Reverse	Phage Tail Protein X
9046	8156	Reverse	Phage P2 GpU
11588	9048	Reverse	Phage-related minor tail protein
12968	12444	Reverse	Phage tail tube protein FII
14404	12971	Reverse	Phage tail sheath protein
15114	14650	Reverse	Gp37 protein
18804	18277	Reverse	Phage virion morphogenesis family protein
19637	18801	Reverse	Phage Mu protein F like protein
22604	21102	Reverse	Terminase-like family protein
25019	24408	Reverse	membrane-bound lytic transglycosylase F
27698	27282	Reverse	Helix-turn-helix domain protein
28950	29252	Forward	IclR helix-turn-helix domain protein
30457	32256	Forward	Integrase core domain protein
35060	35332	Forward	DNA-binding protein HU

36188	36559	Forward	Mor transcription activator family protein

Table 7: Functional genes identified for vB_Bcc_7a.

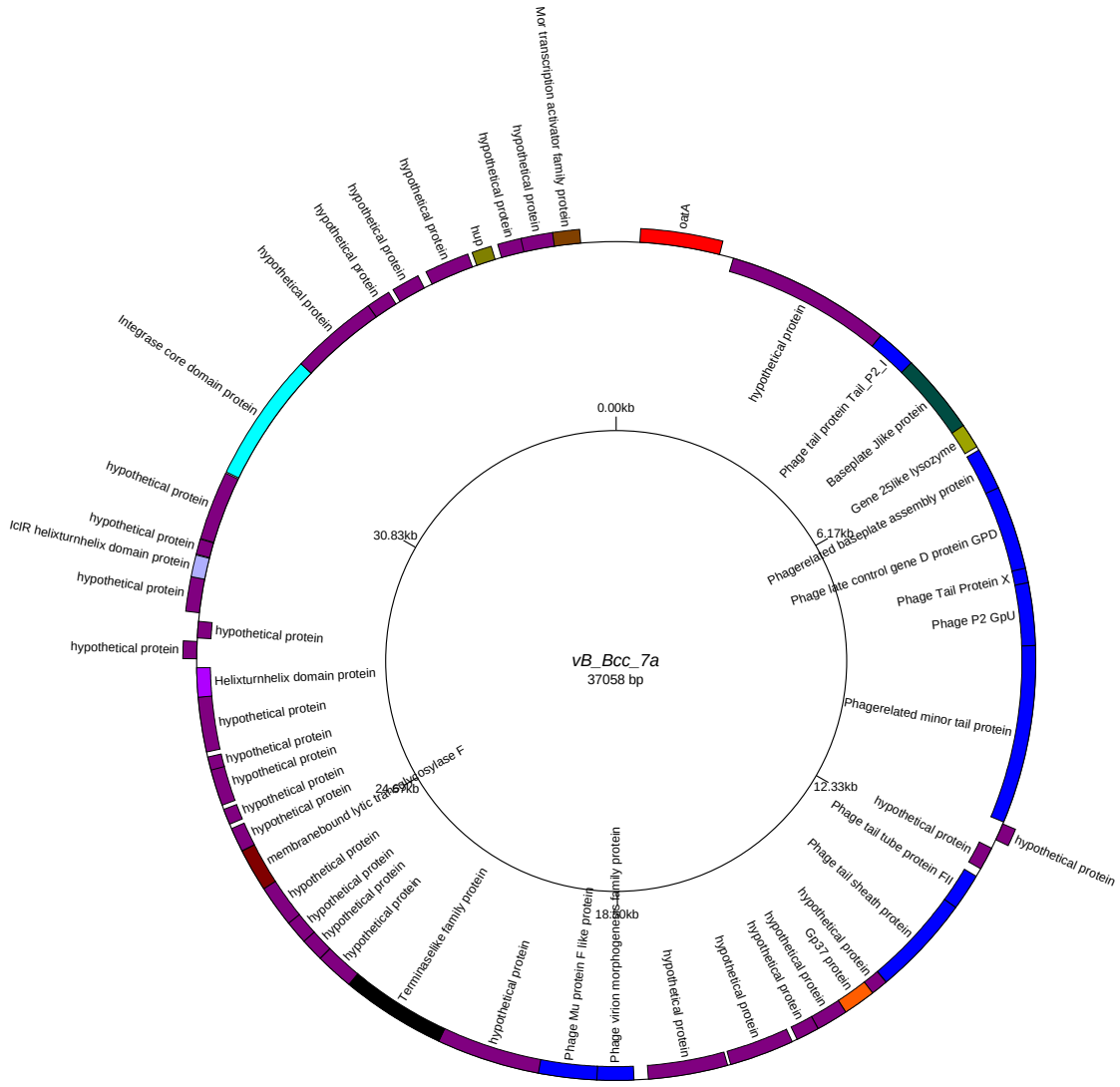


Figure 7: Genome map of vB_Bcc_7a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_11a genome size was 44.9 kb, the phage was assembled using SPAdes and has a mean coverage of 337. The genome map of this phage is shown in Figure 8. The total number of genes identified was 73 of which 18 were identified with putative functions these can be seen in Table 8. Blastn on viruses only database showed a 10.3 kb / 11.1 kb match with *Burkholderia* phage Bcep176 (GenBank: NC_007497.1). The phage vB_Bcc_11a had a 61.9 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
2697	3800	Forward	Modification methylase DpnIIB
7866	9047	Forward	Tyrosine recombinase XerD
12832	12335	Reverse	Lysozyme RrrD
17964	17401	Reverse	Bacteriophage lambda tail assembly protein I
18713	17961	Reverse	NlpC/P60 family protein
19442	18762	Reverse	Phage minor tail protein L
20922	20584	Reverse	Phage minor tail protein
25064	20922	Reverse	Chromosome partition protein Smc
25816	25352	Reverse	Phage tail assembly chaperone
27452	27123	Reverse	Phage head-tail joining protein
27796	27455	Reverse	Phage gp6-like head-tail connector protein
29315	28038	Reverse	Phage capsid family protein
30135	29320	Reverse	ATP-dependent Clp protease proteolytic subunit
31408	30113	Reverse	Phage portal protein
33128	31416	Reverse	Phage Terminase
33607	33128	Reverse	Terminase small subunit
40106	39261	Reverse	Chromosome-partitioning protein Spo0J
41810	42154	Forward	transcriptional repressor DicA

Table 8: Functional genes identified for vB_Bcc_11a.

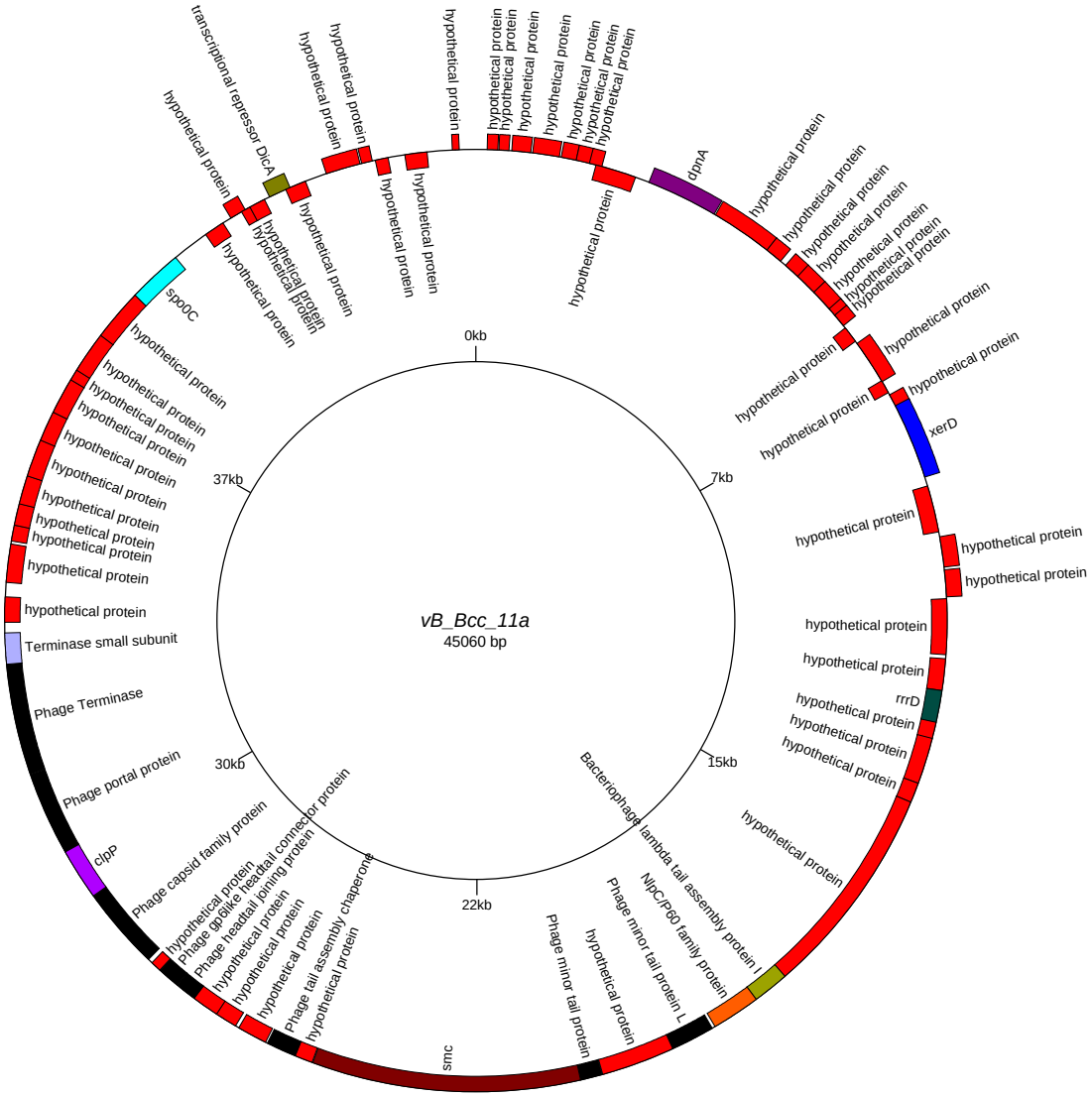


Figure 8: Genome map of vB_Bcc_11a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_11b genome size was 31.8 kb, the phage was assembled using SPAdes and has a mean coverage of 29. The genome map of this phage is shown in Figure 9. The total number of genes identified was 39 of which 25 were identified with putative functions these can be seen in Table 9. Blastn on viruses only database showed an 8.4 kb / 8.5 kb match with *Burkholderia* phage KS5 (GenBank: NC_015265.1). The phage vB_Bcc_11b had a 63.8 % G-C content. A tRNA gene was identified along with several genes associated with structural and regulatory proteins.

Start	End	Direction	Putative functional protein
308	718	Forward	P2 phage tail completion protein R (GpR)
718	1167	Forward	Phage virion morphogenesis family protein
1267	1899	Forward	Phage-related baseplate assembly protein
1896	2273	Forward	Gene 25-like lysozyme
2270	3175	Forward	Baseplate J-like protein
3168	3722	Forward	Phage tail protein (Tail_P2_I)
3725	5335	Forward	Phage Tail Collar Domain protein
5793	6179	Forward	Caudovirales tail fibre assembly protein
6940	7689	Forward	Modification methylase DpnIIB
7801	8973	Forward	Phage tail sheath protein
9003	9512	Forward	Phage tail tube protein FII
9545	9856	Forward	Phage tail protein E
9856	9975	Forward	Phage P2 GpE
12748	13176	Forward	Phage P2 GpU
13173	14318	Forward	Phage late control gene D protein (GPD)
14856	14404	Reverse	tRNA_anti-like protein
16784	17032	Forward	DNA-binding transcriptional regulator
22113	23219	Forward	Tyrosine recombinase XerC
24120	24443	Forward	Transcriptional regulator ClgR

24400	25419	Forward	Reverse transcriptase (RNA-dependent DNA polymerase)
26941	25889	Reverse	Phage portal protein
28611	26941	Reverse	Terminase-like family protein
28855	29676	Forward	Phage capsid scaffolding protein (GPO) serine peptidase
29713	31422	Forward	Phage major capsid protein, P2 family
31526	31855	Forward	Phage head completion protein (GPL)

Table 9: Functional genes identified for vB_Bcc_11b.

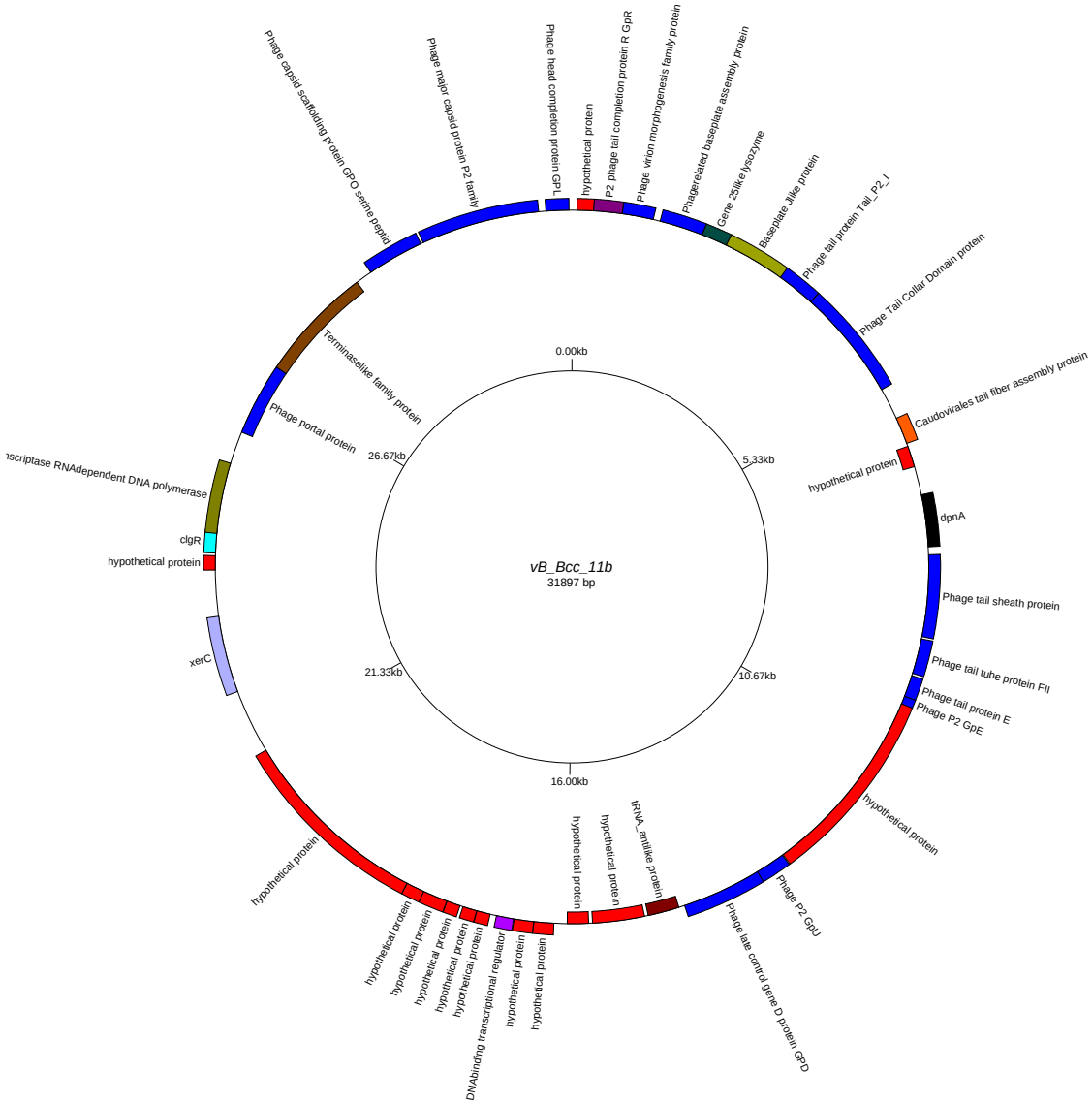
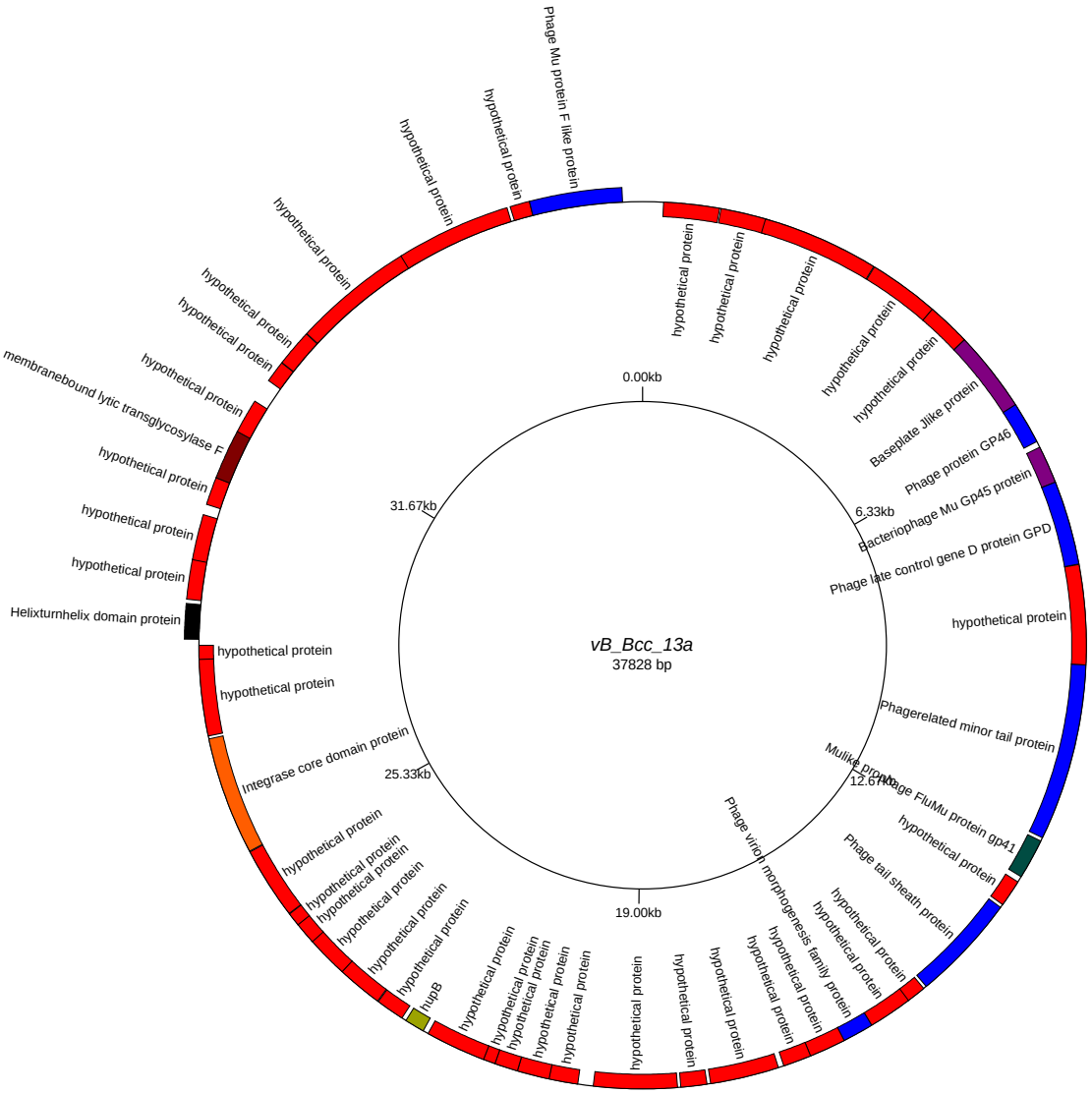


Figure 9: Genome map of vB_Bcc_11b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_13a genome size was 37.8 kb, the phage was assembled using SPAdes and has a mean coverage of 48. The genome map of this phage is shown in Figure 10. The total number of genes identified was 49 of which 14 were identified with putative functions these can be seen in Table 10. Blastn on viruses only database showed a 37.6 kb / 37.6 kb match with *Burkholderia* phage KS10 (GenBank: NC_011216.1). The phage vB_Bcc_13a had a 62.9 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified including an integrase gene.

Start	End	Direction	Putative functional protein
6007	4886	Reverse	Baseplate J-like protein
6585	6004	Reverse	Phage protein GP46
7191	6670	Reverse	Bacteriophage Mu Gp45 protein
8348	7191	Reverse	Phage late control gene D protein (GPD)
9724	8354	Reverse	DNA circulation protein N-terminus
12159	9724	Reverse	Phage-related minor tail protein
12762	12208	Reverse	Mu-like prophage FluMu protein gp41
14740	13262	Reverse	Phage tail sheath protein
16079	15648	Reverse	Phage virion morphogenesis family protein
22302	22030	Reverse	DNA-binding protein HU-beta
27090	25465	Reverse	Integrase core domain protein
28452	28922	Forward	Helix-turn-helix domain protein
30624	31295	Forward	membrane-bound lytic transglycosylase F
36324	37556	Forward	Phage Mu protein F like protein

Table 10: Functional genes identified for vB_Bcc_13a.


vB_Bcc_13a
37828 bp
0.00kb
6.33kb
12.67kb
19.00kb
25.33kb
31.67kb
Phage Mu protein F like protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Baseplate Jlike protein
Phage protein GP46
Bacteriophage Mu Gp45 protein
Phage late control gene D protein GPD
hypothetical protein
Phagerelated minor tail protein
Mulike prophage FluMu protein gp41
hypothetical protein
Phage tail sheath protein
hypothetical protein
Phage virion morphogenesis family protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The potentially incomplete temperate phage vB_Bcc_15a genome size was 29.6 kb, the phage was assembled using SPAdes and has a mean coverage of 122. The genome map of this phage is shown in Figure 11. The total number of genes identified was 45 of which 25 were identified with putative functions these can be seen in Table 11. Blastn on viruses only database showed a 2.6 kb / 3.3 kb match with *Burkholderia* phage phiE12-2 (GenBank: NC_009236.1). The phage vB_Bcc_15a had a 62.7 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified including an integrase gene.

Start	End	Direction	Putative functional protein
1218	562	Reverse	Chromosome-partitioning ATPase Soj
1511	1305	Reverse	Ogr/Delta-like zinc finger
1774	1508	Reverse	Ogr/Delta-like zinc finger
3589	1787	Reverse	Bacteriophage replication gene A protein (GPA)
6048	6377	Forward	Helix-turn-helix
9024	7960	Reverse	Phage late control gene D protein (GPD)
9731	9021	Reverse	Phage P2 GpU
12525	12412	Reverse	Phage P2 GpE
12914	12534	Reverse	Phage tail protein E
13495	12986	Reverse	Phage tail tube protein FII
14708	13533	Reverse	Phage tail sheath protein
16808	16245	Reverse	Phage tail protein (Tail_P2_I)
17712	16801	Reverse	Baseplate J-like protein
18080	17709	Reverse	Gene 25-like lysozyme
18766	18077	Reverse	Phage-related baseplate assembly protein
19315	18845	Reverse	Phage virion morphogenesis family protein
19749	19300	Reverse	P2 phage tail completion protein R (GpR)
21118	20312	Reverse	Putative peptidoglycan binding domain protein
21963	21757	Reverse	Phage Tail Protein X
22795	22313	Reverse	Phage head completion protein (GPL)
23617	22898	Reverse	Phage small terminase subunit
24639	23620	Reverse	Phage major capsid protein, P2 family

25741	24695	Reverse	Phage capsid scaffolding protein (GPO) serine peptidase
25783	27666	Forward	Terminase-like family protein
27663	28715	Forward	Phage portal protein

Table 11: Functional genes identified for vB_Bcc_15a.

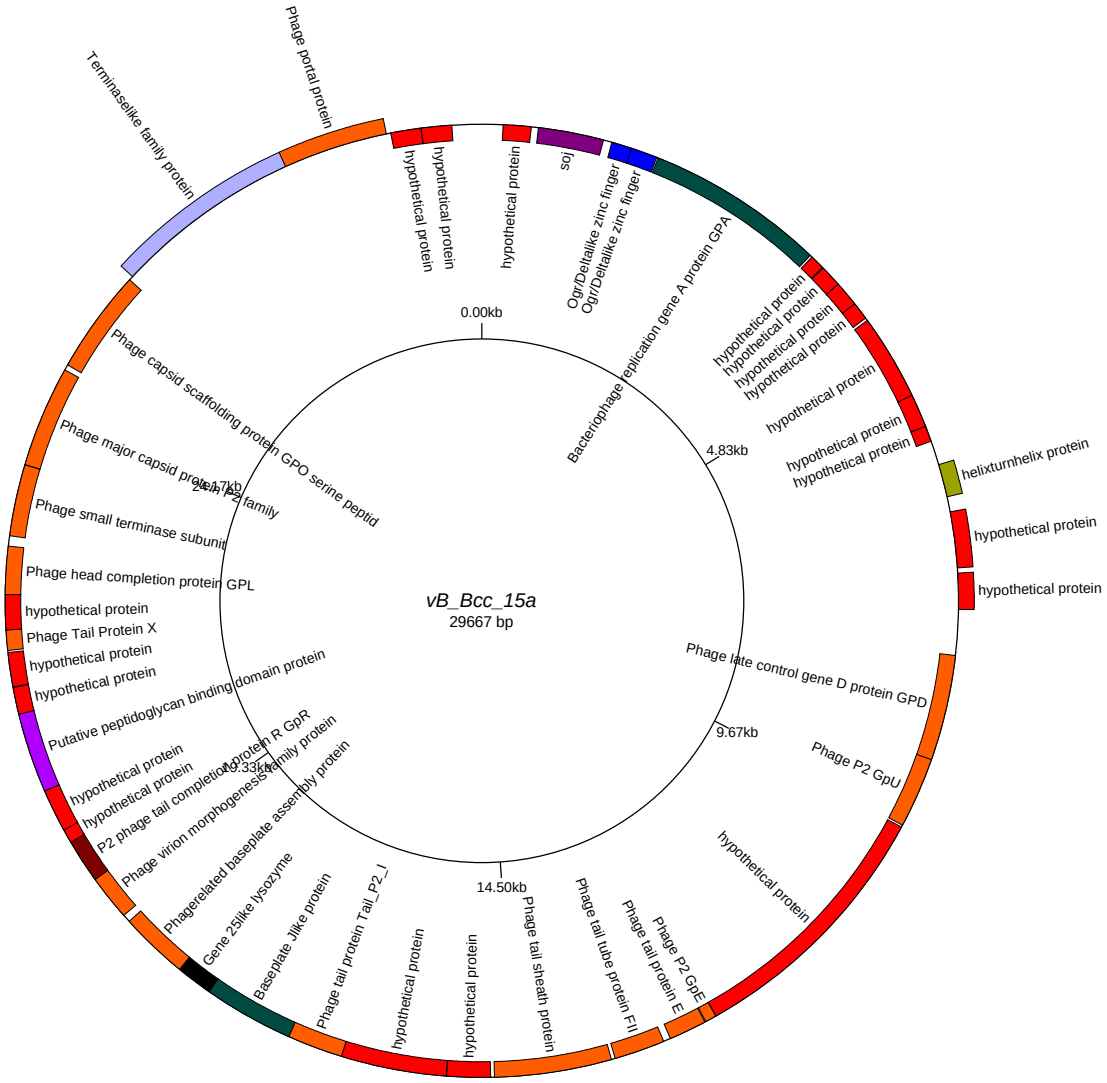


Figure 11: Genome map of vB_Bcc_15a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_16a genome size was 37.2 kb, the phage was assembled using SPAdes and has a mean coverage of 36. The genome map of this phage is shown in Figure 12. The total number of genes identified was 45 of which 22 were identified with putative functions these can be seen in Table 12. Blastn on viruses only database showed a 786 bp / 949 bp match with *Burkholderia* phage KS9 (GenBank: NC_013055.1). The phage vB_Bcc_16a had a 62.4 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
3471	976	Reverse	DNA primase TraC
9866	10120	Forward	Prophage CP4-57 regulatory protein (AlpA)
13413	12109	Reverse	Phage integrase family protein
16039	15251	Reverse	Modification methylase DpnIIA
17194	16700	Reverse	Phage lysozyme
17481	17197	Reverse	Phage holin family 2
18608	17556	Reverse	Phage late control gene D protein (GPD)
18825	18619	Reverse	Phage Tail Protein X
19681	18800	Reverse	Phage P2 GpU
22109	19692	Reverse	Phage-related minor tail protein
23080	22577	Reverse	Phage tail tube protein FII
24260	23091	Reverse	Phage tail sheath protein
25067	24357	Reverse	Caudovirales tail fibre assembly protein
26384	25080	Reverse	Phage Tail Collar Domain protein
26950	26381	Reverse	Phage tail protein (Tail_P2_I)
27833	26940	Reverse	Baseplate J-like protein
28174	27830	Reverse	Gene 25-like lysozyme
29121	28441	Reverse	Phage-related baseplate assembly protein
30165	29638	Reverse	Prophage minor tail protein Z (GPZ)
32959	31904	Reverse	ATP-dependent Clp protease proteolytic subunit
34446	32956	Reverse	Phage portal protein, lambda family
36647	34659	Reverse	Phage terminase large subunit (GpA)

Table 12: Functional genes identified for vB_Bcc_16a.

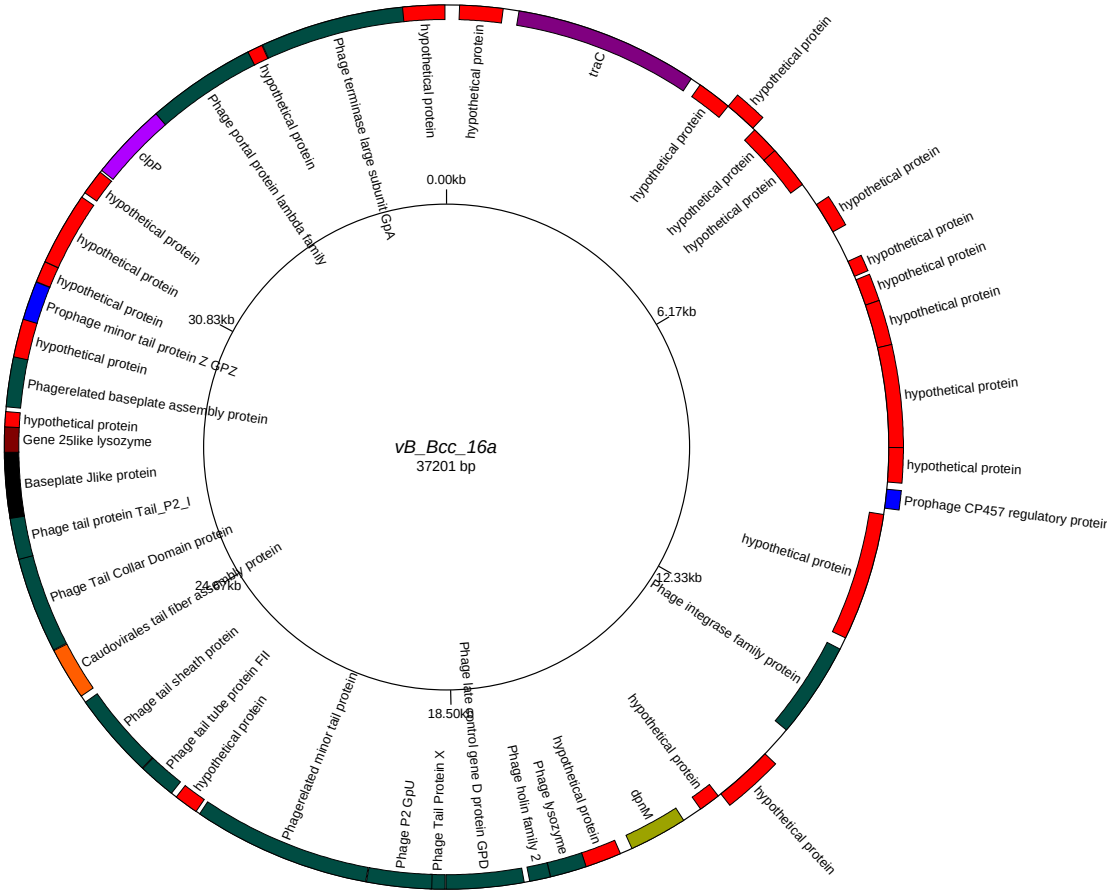


Figure 12: Genome map of vB_Bcc_16a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_17a genome size was 35.8 kb, the phage was assembled using SPAdes and has a mean coverage of 37. The genome map of this phage is shown in Figure 13. The total number of genes identified was 45 of which 12 were identified with putative functions these can be seen in Table 13. Blastn on viruses only database showed no matches. Thus a Blastn search on the remote database showed a 35.8 kb / 35.8 kb match to *Burkholderia cenocepacia* J2315 chromosome 2 (GenBank: AM747721.1). The phage vB_Bcc_17a had a 64.8 % G-C content. No tRNA gene was identified however; several genes associated regulatory proteins were identified.

Start	End	Direction	Putative functional protein
2835	1294	Reverse	Glycerol kinase
3116	2832	Reverse	Major Facilitator Superfamily protein
4295	3171	Reverse	putative metabolite transport protein CsbC
5255	4536	Reverse	C-factor
5388	6365	Forward	HTH-type transcriptional regulator SyrM 1
6480	7868	Forward	Xanthine permease XanP
8000	8425	Forward	RNA pyrophosphohydrolase
9715	9119	Reverse	Lipase (class 3)
11676	11104	Reverse	Chitinase class I
14073	12517	Reverse	Pectate lyase superfamily protein
23561	21774	Reverse	Peptidoglycan hydrolase FlgJ
35399	33852	Reverse	Terminase-like family protein

Table 13: Functional genes identified for vB_Bcc_17a.

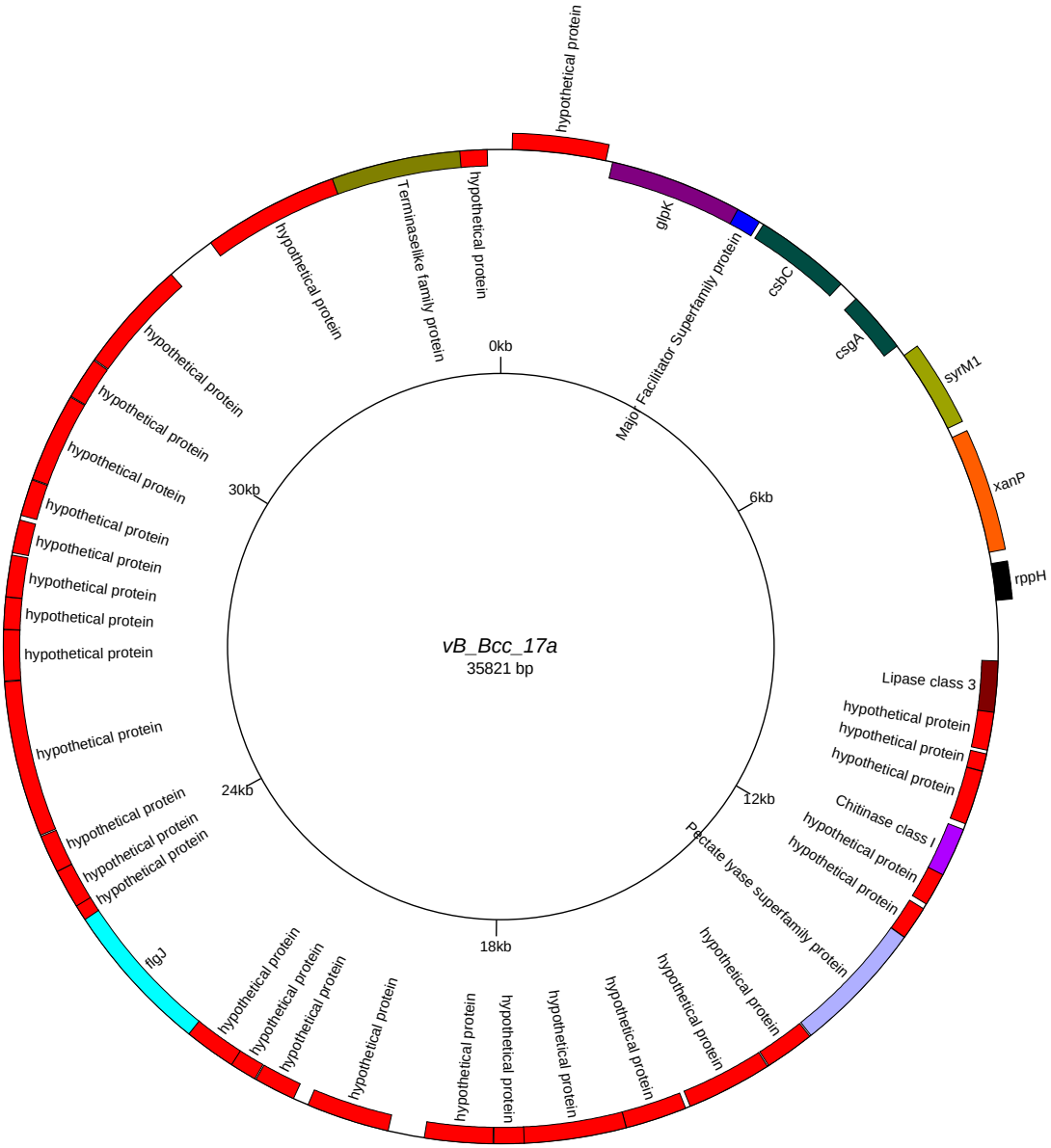


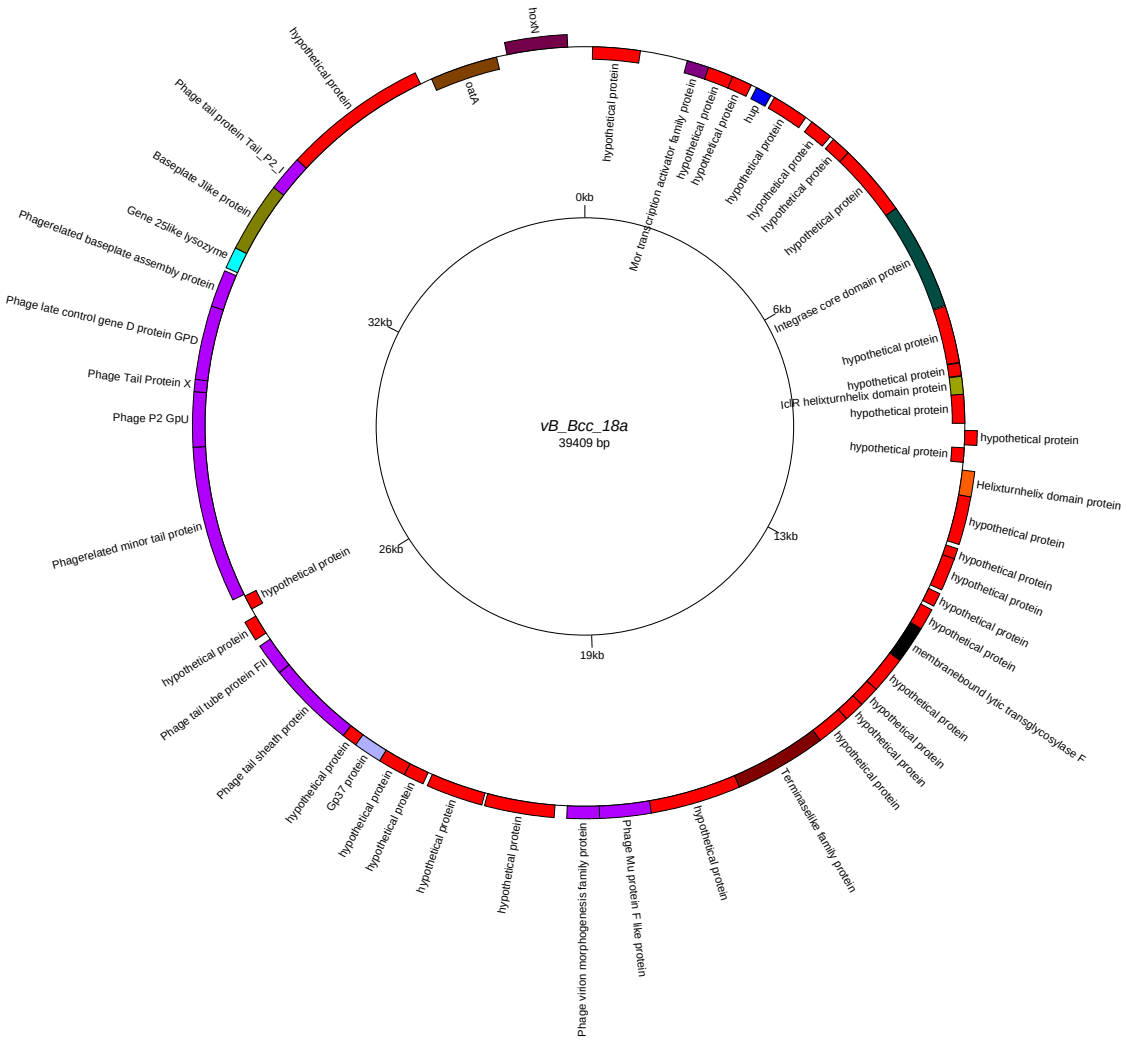
Figure 13: Genome map of vB_Bcc_17a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_18a genome size was 39.4 kb, the phage was assembled using SPAdes and has a mean coverage of 855. The genome map of this phage is shown in Figure 14. The total number of genes identified was 52 of which 22 were identified with putative functions these can be seen in Table 14. Blastn on viruses only database showed a 36.7 kb / 36.7 kb match with *Burkholderia* phage BcepMu (GenBank: NC_005882.1). The phage vB_Bcc_18a had a 63.2 % G-C content. No tRNA gene was identified however; several genes associated with structural, lysogenic and regulatory proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
2084	1713	Reverse	Mor transcription activator family protein
3212	2940	Reverse	DNA-binding protein HU
7815	6016	Reverse	Integrase core domain protein
9322	9020	Reverse	IclR helix-turn-helix domain protein
10574	10990	Forward	Helix-turn-helix domain protein
13253	13864	Forward	membrane-bound lytic transglycosylase F
15668	17170	Forward	Terminase-like family protein
18635	19471	Forward	Phage Mu protein F like protein
19468	19995	Forward	Phage virion morphogenesis family protein
23158	23622	Forward	Gp37 protein
23868	25301	Forward	Phage tail sheath protein
25304	25828	Forward	Phage tail tube protein FII
26684	29224	Forward	Phage-related minor tail protein
29226	30116	Forward	Phage P2 GpU
30116	30325	Forward	Phage Tail Protein X
30313	31518	Forward	Phage late control gene D protein (GPD)
31515	32117	Forward	Phage-related baseplate assembly protein
32171	32524	Forward	Gene 25-like lysozyme
32521	33672	Forward	Baseplate J-like protein
33665	34246	Forward	Phage tail protein (Tail_P2_I)

37937	36804	Reverse	O-acetyltransferase OatA
38115	39125	Forward	High-affinity nickel transport protein

Table 14: Functional genes identified for vB_Bcc_18a.


vB_Bcc_18a
39409 bp
0kb
6kb
13kb
19kb
26kb
32kb
NoxH
hypothetical protein
gatA
Phage tail protein Tail_P2_I
Baseplate Jlike protein
Gene 25like lysozyme
Phagerelated baseplate assembly protein
Phage late control gene D protein GPD
Phage Tail Protein X
Phage P2 GpU
Phagerelated minor tail protein
hypothetical protein
hypothetical protein
Phage tube protein FII
Phage tail sheath protein
hypothetical protein
Gp37 protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage virion morphogenesis family protein
Phage Mu protein F like protein
hypothetical protein
Terminaselike family protein
hypothetical protein
hypothetical protein
hypothetical protein
membranebound lytic transglycosylase F
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Helixturnhelix domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
IcoR helixturnhelix domain protein
hypothetical protein
Integrase core domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical

The temperate phage vB_Bcc_19a genome size was 36.9 kb, the phage was assembled using SPAdes and has a mean coverage of 806. The genome map of this phage is shown in Figure 15. The total number of genes identified was 52 of which 22 were identified with putative functions these can be seen in Table 15. Blastn on viruses only database showed a 36.7 kb / 36.7 kb match with *Burkholderia* phage BcepMu (GenBank: NC_005882.1). The phage vB_Bcc_19a had a 62.9 % G-C content. No tRNA gene was identified however; several genes associated with structural, lysogenic and regulatory proteins were identified.

Start	**End**	**Direction**	**Putative functional protein**
829	458	Reverse	Mor transcription activator family protein
1957	1685	Reverse	DNA-binding protein HU
6560	4761	Reverse	Integrase core domain protein
8067	7765	Reverse	IclR helix-turn-helix domain protein
9319	9735	Forward	Helix-turn-helix domain protein
11998	12609	Forward	membrane-bound lytic transglycosylase F
14413	15915	Forward	Terminase-like family protein
17380	18216	Forward	Phage Mu protein F like protein
18213	18740	Forward	Phage virion morphogenesis family protein
21903	22367	Forward	Gp37 protein
22613	24046	Forward	Phage tail sheath protein
24049	24573	Forward	Phage tail tube protein FII
25429	27969	Forward	Phage-related minor tail protein
27971	28861	Forward	Phage P2 GpU
28861	29070	Forward	Phage Tail Protein X
29058	30263	Forward	Phage late control gene D protein (GPD)
30260	30862	Forward	Phage-related baseplate assembly protein
30916	31269	Forward	Gene 25-like lysozyme
31266	32417	Forward	Baseplate J-like protein
32410	32991	Forward	Phage tail protein (Tail_P2_I)

36682	35549	Reverse	O-acetyltransferase OatA

Table 15: Functional genes identified for vB_Bcc_19a.

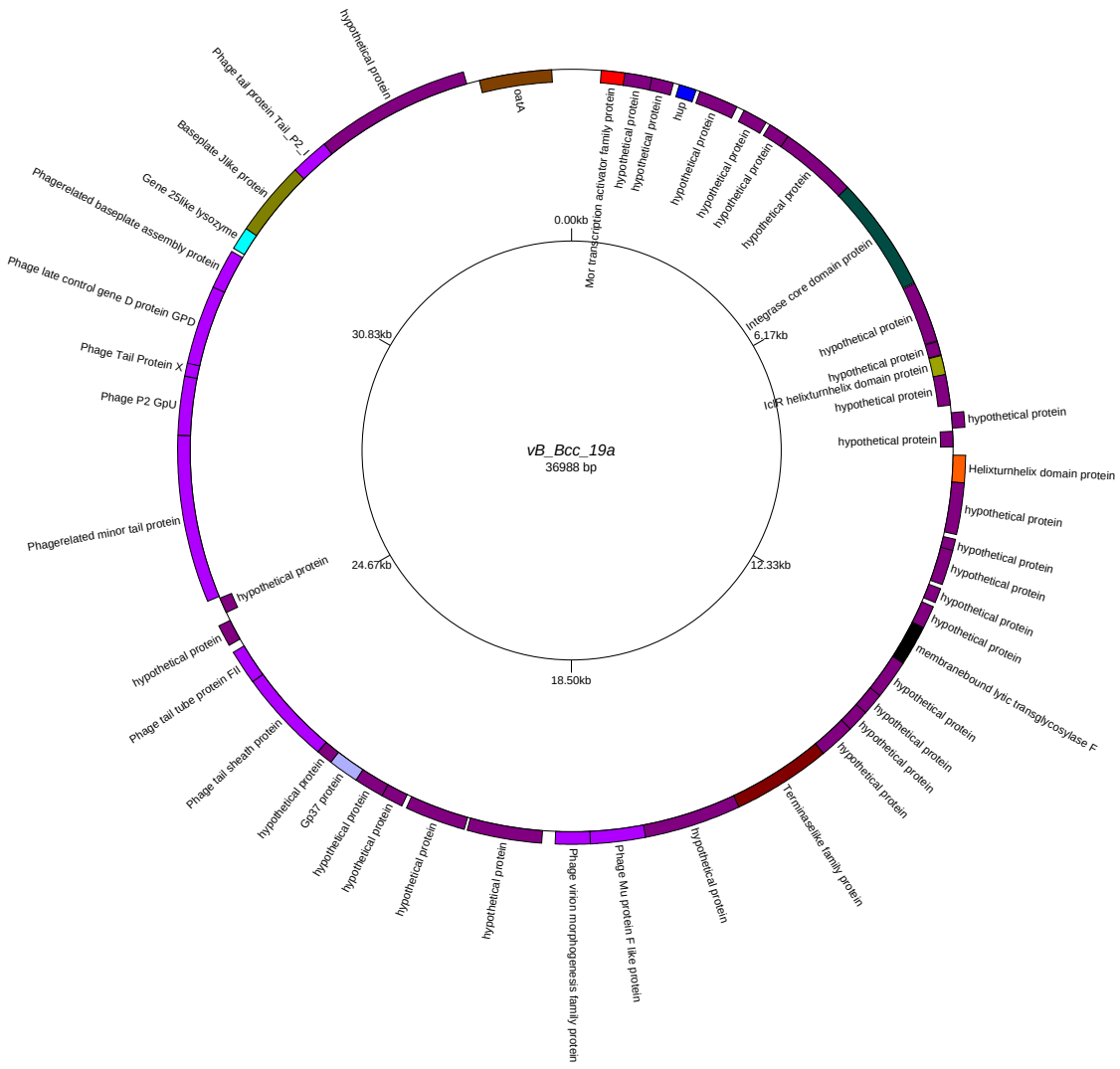


Figure 15: Genome map of vB_Bcc_19a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_23a genome size was 40.7 kb, the phage was assembled using SPAdes and has a mean coverage of 57. The genome map of this phage is shown in Figure 16. The total number of genes identified was 54 of which 20 were identified with putative functions these can be seen in Table 16. Blastn on viruses only database showed an 8.2 kb / 8.4 kb match with *Burkholderia* phage Bcep176 (GenBank: NC_007497.1). The phage vB_Bcc_23a had a 60.8 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
632	2341	Forward	Phage Terminase
2338	3627	Forward	Phage portal protein
3593	4390	Forward	ATP-dependent Clp protease proteolytic subunit
4458	5720	Forward	Phage capsid family protein
5770	6354	Forward	Phage gp6-like head-tail connector protein
6365	6691	Forward	Phage head-tail joining protein
7997	8461	Forward	Phage tail assembly chaperone
8749	12891	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
12891	13229	Forward	Phage minor tail protein
14137	14820	Forward	Phage minor tail protein L
14869	15621	Forward	NlpC/P60 family protein
15618	16181	Forward	Bacteriophage lambda tail assembly protein I
20750	21247	Forward	Lysozyme RrrD
27030	28304	Forward	integrase
28530	28249	Reverse	Prophage CP4-57 regulatory protein (AlpA)
29378	28527	Reverse	Chromosome-partitioning protein Spo0J
30166	29387	Reverse	thiamine biosynthesis protein ThiF
35929	35465	Reverse	Helix-turn-helix
37044	38414	Forward	Replicative DNA helicase
40357	40695	Forward	HNH endonuclease

Table 16: Functional genes identified for vB_Bcc_23a.

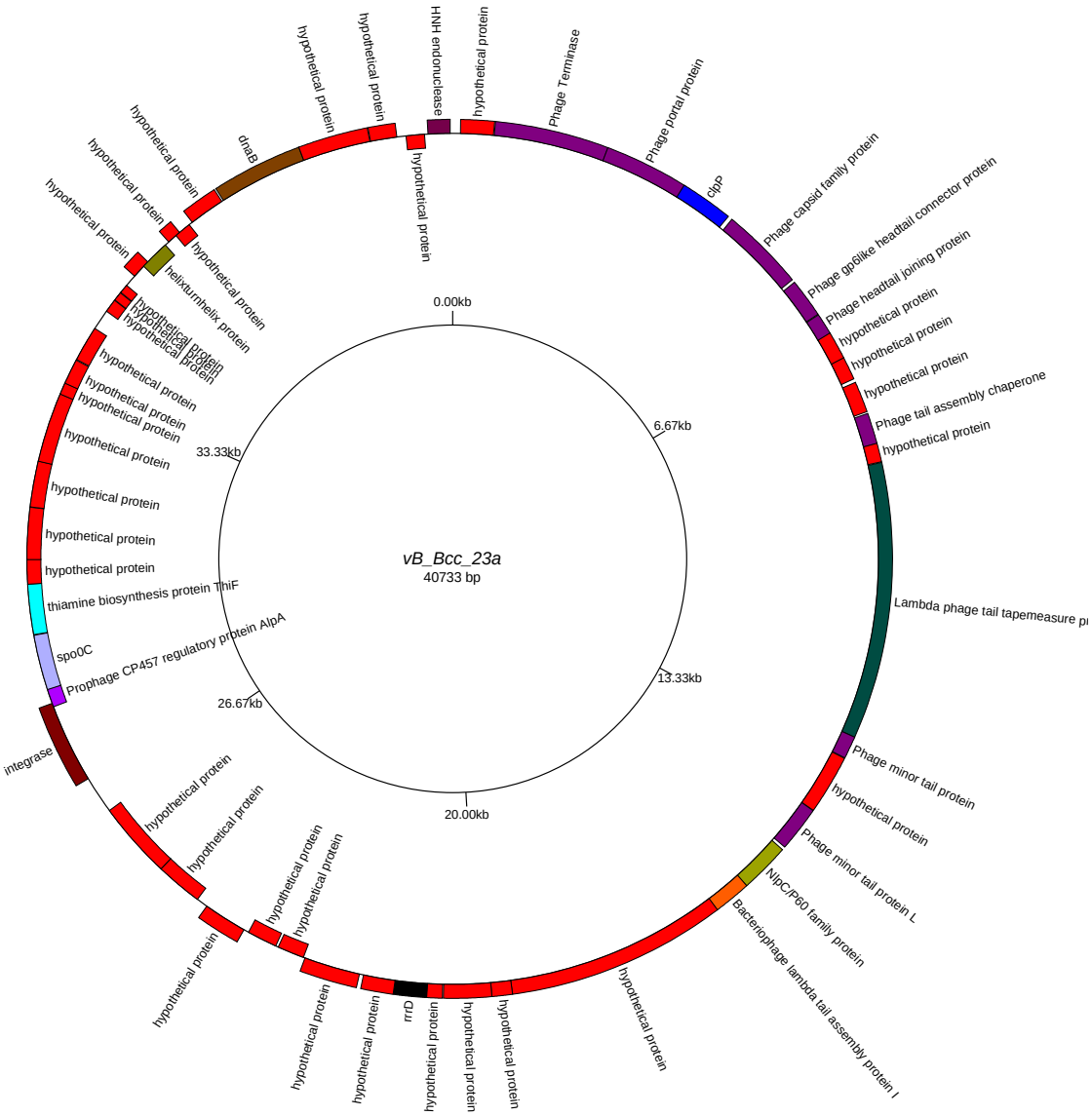


Figure 16: Genome map of vB_Bcc_23a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_28a genome size was 44.7 kb, the phage was assembled using SPAdes and has a mean coverage of 16. The genome map of this phage is shown in Figure 17. The total number of genes identified was 73 of which 11 were identified with putative functions these can be seen in Table 17. Blastn on viruses only database showed a 127 bp / 154 bp match with *Burkholderia* phage Bcep176 (GenBank: NC_007497.1). The phage vB_Bcc_28a had a 63.2 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
3872	3489	Reverse	Fic/DOC family protein
7556	8026	Forward	Endodeoxyribonuclease RusA
9590	11137	Forward	Terminase-like family protein
12656	13297	Forward	Phage Mu protein F like protein
21428	23215	Forward	Peptidoglycan hydrolase FlgJ
30916	32475	Forward	Pectate lyase superfamily protein
33320	33886	Forward	Chitinase class I
35275	35871	Forward	Lipase (class 3)
37158	36169	Reverse	site-specific tyrosine recombinase XerC
39850	39608	Reverse	Helix-turn-helix domain protein
43472	42792	Reverse	YqaJ-like viral recombinase domain protein

Table 17: Functional genes identified for vB_Bcc_28a.

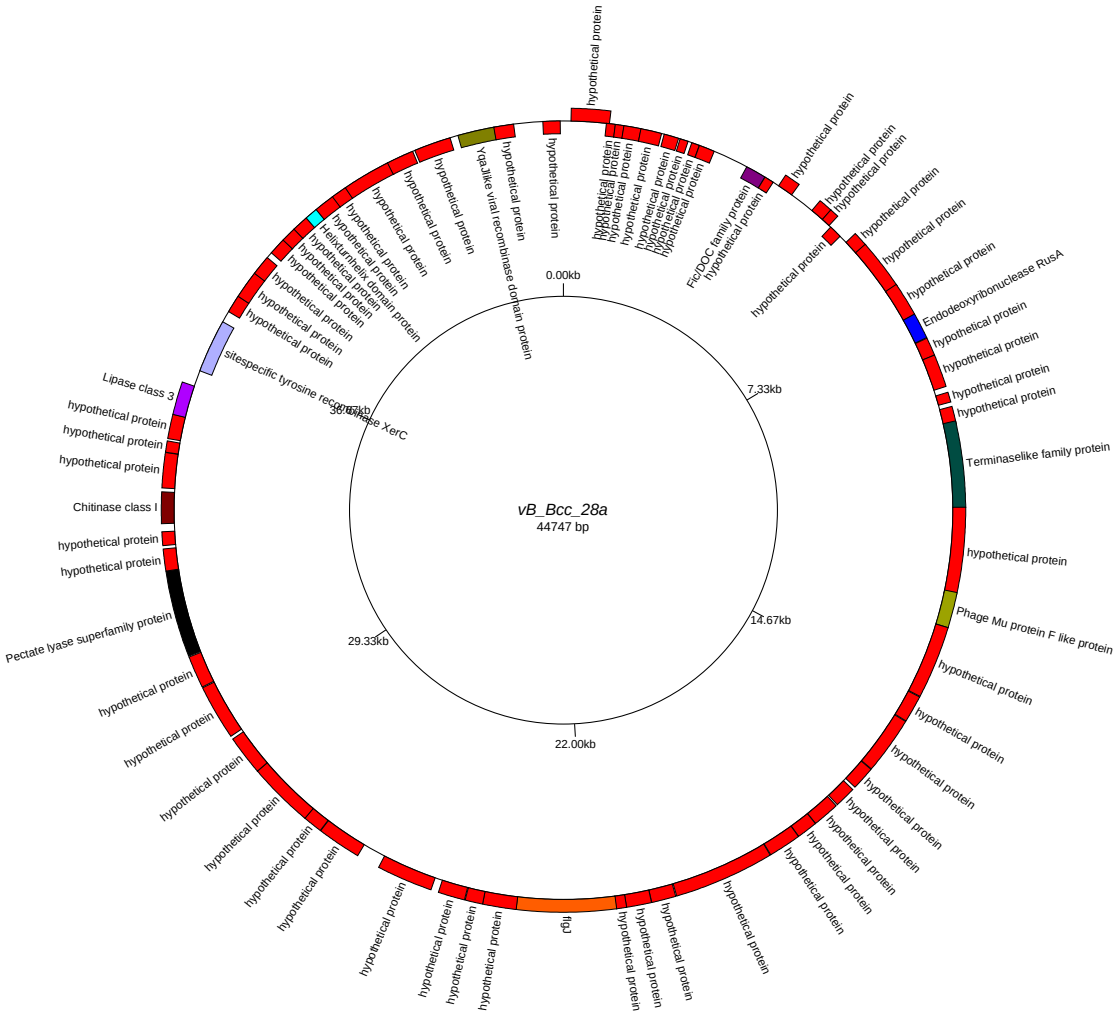


Figure 17: Genome map of vB_Bcc_28a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_28b genome size was 37.7 kb, the phage was assembled using SPAdes and has a mean coverage of 92. The genome map of this phage is shown in Figure 18. The total number of genes identified was 50 of which 14 were identified with putative functions these can be seen in Table 18. Blastn on viruses only database showed a 37.6 kb / 37.6 kb match with *Burkholderia* phage KS10 (GenBank: NC_011216.1). The phage vB_Bcc_28b had a 62.9% G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified, including an integrase gene.

Start	End	Direction	Putative functional protein
1459	227	Reverse	Phage Mu protein F like protein
7159	6488	Reverse	membrane-bound lytic transglycosylase F
9331	8861	Reverse	Helix-turn-helix domain protein
10693	12318	Forward	Integrase core domain protein
15481	15753	Forward	DNA-binding protein HU-beta
21704	22135	Forward	Phage virion morphogenesis family protein
23043	24521	Forward	Phage tail sheath protein
25021	25575	Forward	Mu-like prophage FluMu protein gp41
25624	28059	Forward	Phage-related minor tail protein
28059	29429	Forward	DNA circulation protein N-terminus
29435	30592	Forward	Phage late control gene D protein (GPD)
30592	31113	Forward	Bacteriophage Mu Gp45 protein
31198	31779	Forward	Phage protein GP46
31776	32897	Forward	Baseplate J-like protein

Table 18: Functional genes identified for vB_Bcc_28b.

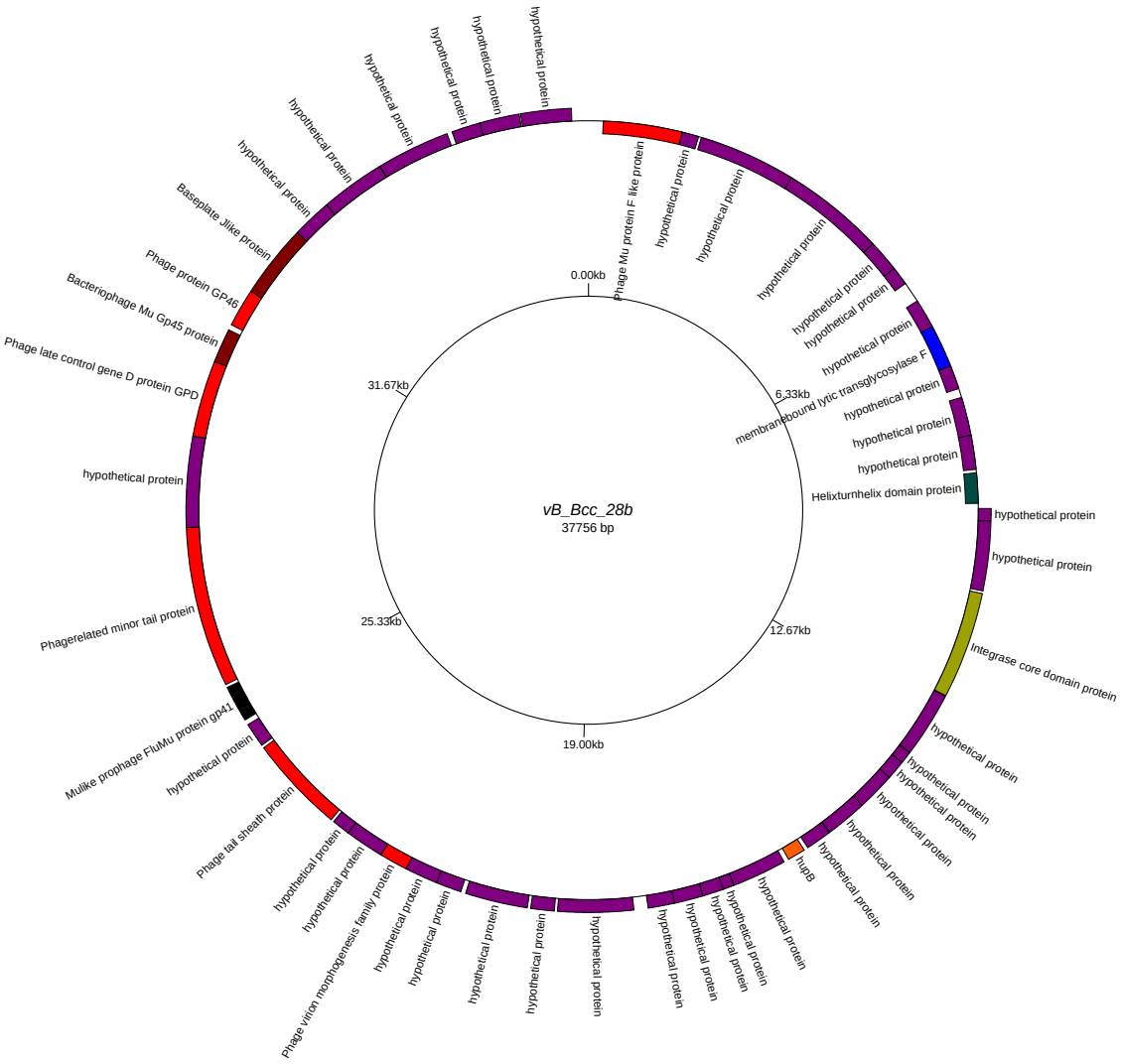
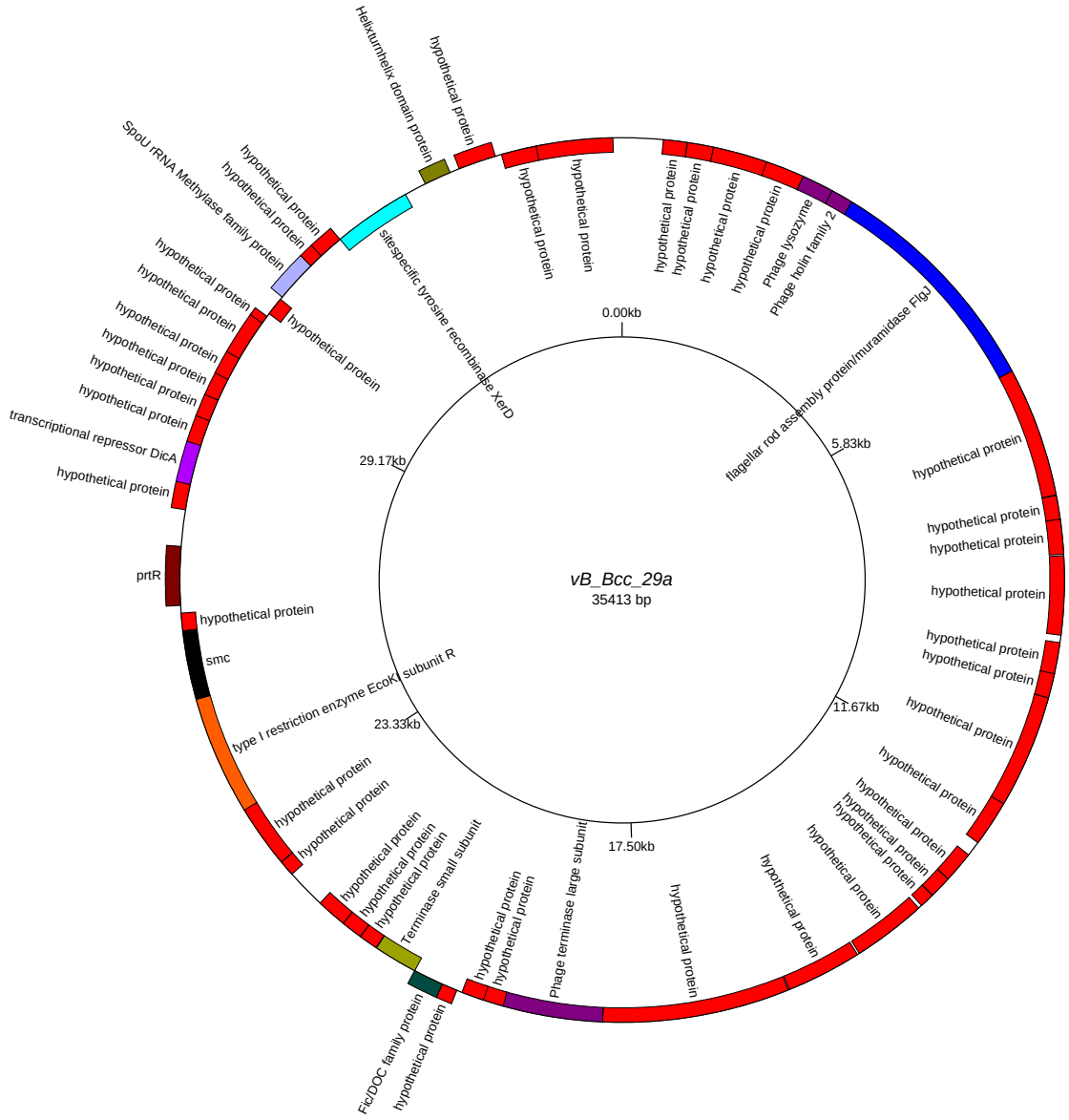


Figure 18: Genome map of vB_Bcc_28b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_29a genome size was 37.7 kb, the phage was assembled using SPAdes and has a mean coverage of 2321. The genome map of this phage is shown in Figure 19. The total number of genes identified was 53 of which 13 were identified with putative functions these can be seen in Table 19. Blastn on viruses only database showed a 307 bp / 346 bp match with *Burkholderia* phage DC1 (GenBank: NC_018452.1). The phage vB_Bcc_29a had a 62.2% G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
2794	2375	Reverse	Phage lysozyme
3078	2791	Reverse	Phage holin family 2
6080	3078	Reverse	flagella rod assembly protein/muramidase FlgJ
19262	17955	Reverse	Phage terminase large subunit
20067	20459	Forward	Fic/DOC family protein
21025	20456	Reverse	Terminase small subunit
25025	23478	Reverse	type I restriction enzyme EcoKI subunit R
25914	25018	Reverse	Chromosome partition protein Smc
26237	26989	Forward	HTH-type transcriptional regulator PrtR
27787	28308	Forward	transcriptional repressor DicA
30472	31014	Forward	SpoU rRNA Methylase family protein
32526	31516	Reverse	site-specific tyrosine recombinase XerD
32814	33161	Forward	Helix-turn-helix domain protein

Table 19: Functional genes identified for vB_Bcc_29a.


vB_Bcc_29a
35413 bp
0.00kb
5.83kb
11.67kb
17.50kb
23.33kb
29.17kb
hypothetical protein
Helixturnhelix domain protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage lysozyme
Phage holin family 2
flagellar rod assembly protein/muramidase FlgJ
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Bcc_31a genome size was 41 kb, the phage was assembled using SPAdes and has a mean coverage of 378. The genome map of this phage is shown in Figure 20. The total number of genes identified was 57 of which 22 were identified with putative functions these can be seen in Table 20. Blastn on viruses only database showed a 770 bp / 913 bp match with *Burkholderia* phage phi1026b (GenBank: NC_005284.1). The phage vB_Bcc_31a had a 61.8% G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
1918	2172	Forward	Prophage CP4-57 regulatory protein (AlpA)
4614	3340	Reverse	Phage integrase family protein
5978	5088	Reverse	HTH-type transcriptional regulator DmlR
6276	6647	Forward	Alkaline phosphatase synthesis transcriptional regulatory protein PhoP
10967	10743	Reverse	Modification methylase DpnIIA
12121	11627	Reverse	Phage lysozyme
12408	12124	Reverse	Phage holin family 2
13536	12484	Reverse	Phage late control gene D protein (GPD)
13753	13547	Reverse	Phage Tail Protein X
14609	13728	Reverse	Phage P2 GpU
17036	14619	Reverse	Phage-related minor tail protein
18018	17515	Reverse	Phage tail tube protein FII
19198	18029	Reverse	Phage tail sheath protein
21361	20795	Reverse	Phage tail protein (Tail_P2_I)
22247	21354	Reverse	Baseplate J-like protein
22588	22244	Reverse	Gene 25-like lysozyme
23536	22856	Reverse	Phage-related baseplate assembly protein
24580	24053	Reverse	Prophage minor tail protein Z (GPZ)
27374	26319	Reverse	ATP-dependent Clp protease proteolytic subunit
28861	27371	Reverse	Phage portal protein, lambda family
31062	29074	Reverse	Phage terminase large subunit (GpA)

36502	34007	Reverse	DNA primase TraC

Table 20: Functional genes identified for vB_Bcc_31a.

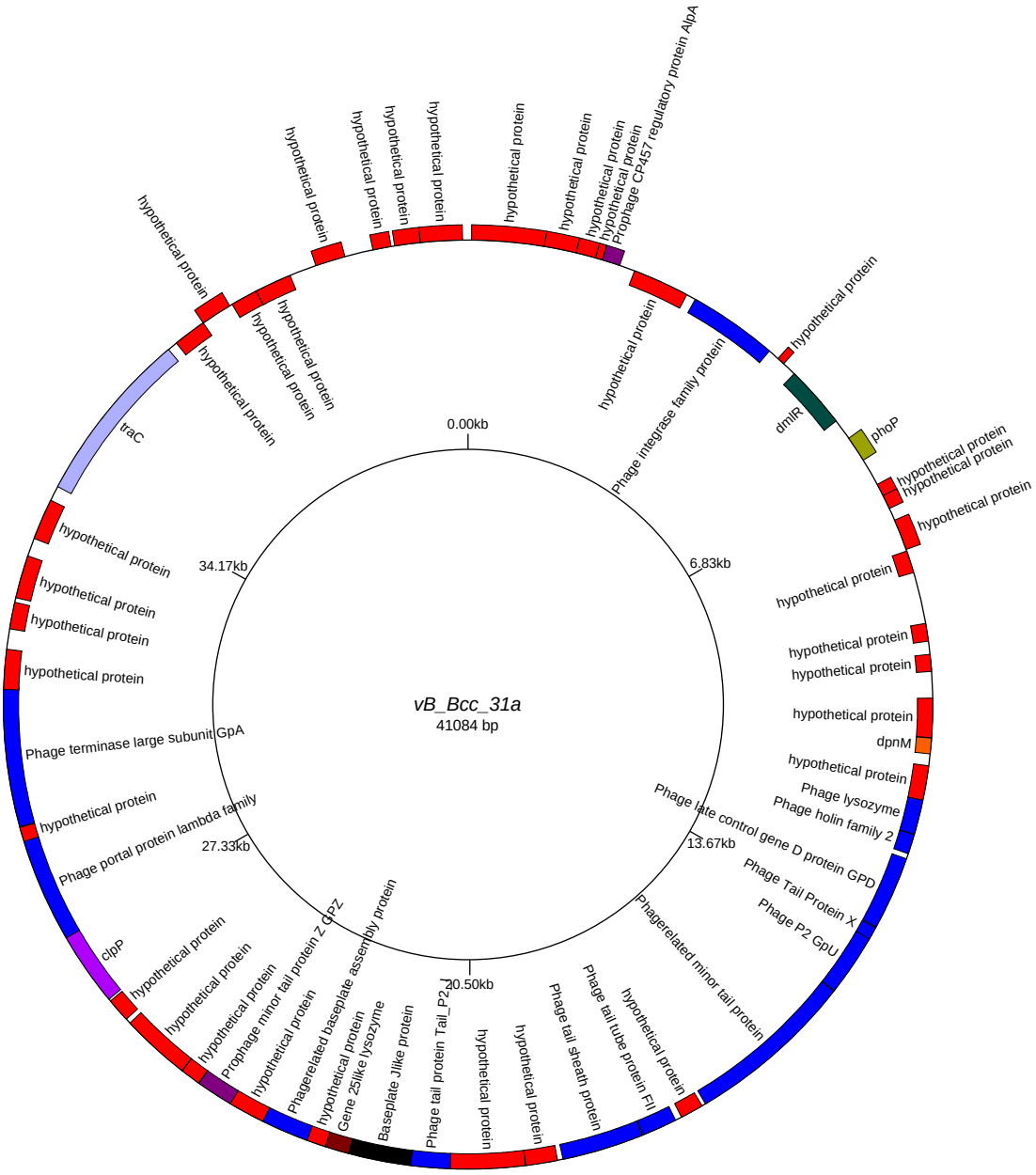


Figure 20: Genome map of vB_Bcc_31a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_32a genome size was 43.6 kb, the phage was assembled using SPAdes and has a mean coverage of 56. The genome map of this phage is shown in Figure 21. The total number of genes identified was 61 of which 24 were identified with putative functions these can be seen in Table 21. Blastn on viruses only database showed a 1048 bp / 1229 bp match with *Ralstonia phage* phiRSA1 (GenBank: NC_009382.1). The phage vB_Bcc_32a had a 61.5% G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified, including an integrase gene.

Start	End	Direction	Putative functional protein
570	2558	Forward	Phage terminase large subunit (GpA)
2771	4261	Forward	Phage portal protein, lambda family
4258	5331	Forward	ATP-dependent Clp protease proteolytic subunit
7067	7594	Forward	Prophage minor tail protein Z (GPZ)
8111	8791	Forward	Phage-related baseplate assembly protein
9058	9402	Forward	Gene 25-like lysozyme
9399	10292	Forward	Baseplate J-like protein
10285	10851	Forward	Phage tail protein (Tail_P2_I)
12477	13646	Forward	Phage tail sheath protein
13657	14160	Forward	Phage tail tube protein FII
14628	17045	Forward	Phage-related minor tail protein
17056	17937	Forward	Phage P2 GpU
17912	18118	Forward	Phage Tail Protein X
18129	19181	Forward	Phage late control gene D protein (GPD)
19321	20541	Forward	Transposase
20569	20853	Forward	Phage holin family 2
20856	21350	Forward	Phage lysozyme
22011	22799	Forward	Modification methylase DpnIIA
25676	26575	Forward	Protease HtpX
27983	29257	Forward	Phage integrase family protein
30001	30333	Forward	DNA-binding transcriptional repressor PuuR

30284	31285	Forward	Reverse transcriptase (RNA-dependent DNA polymerase)
31782	31528	Reverse	Prophage CP4-57 regulatory protein (AlpA)
38809	41304	Forward	DNA primase TraC

Table 21: Functional genes identified for vB_Bcc_32a.

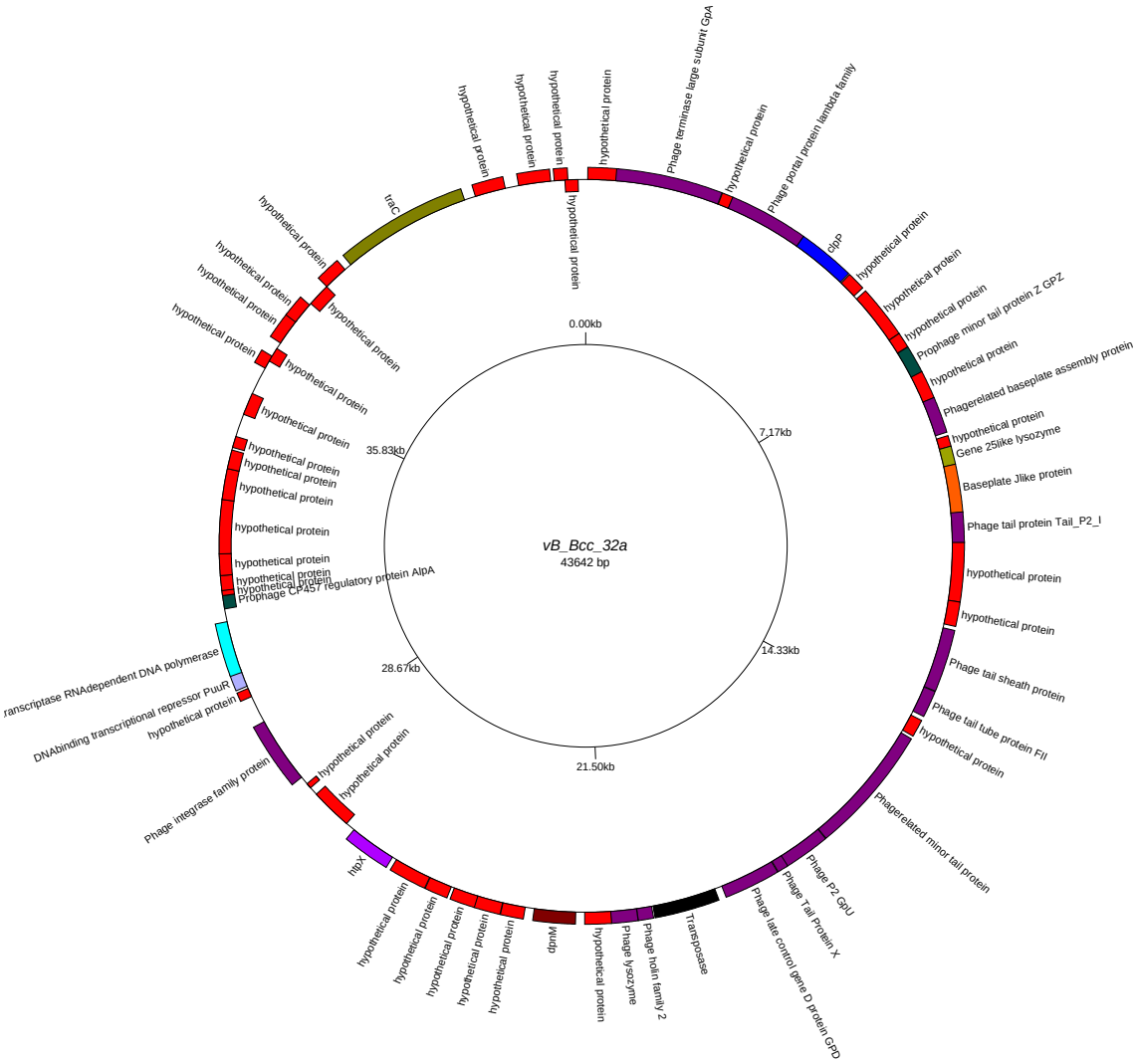
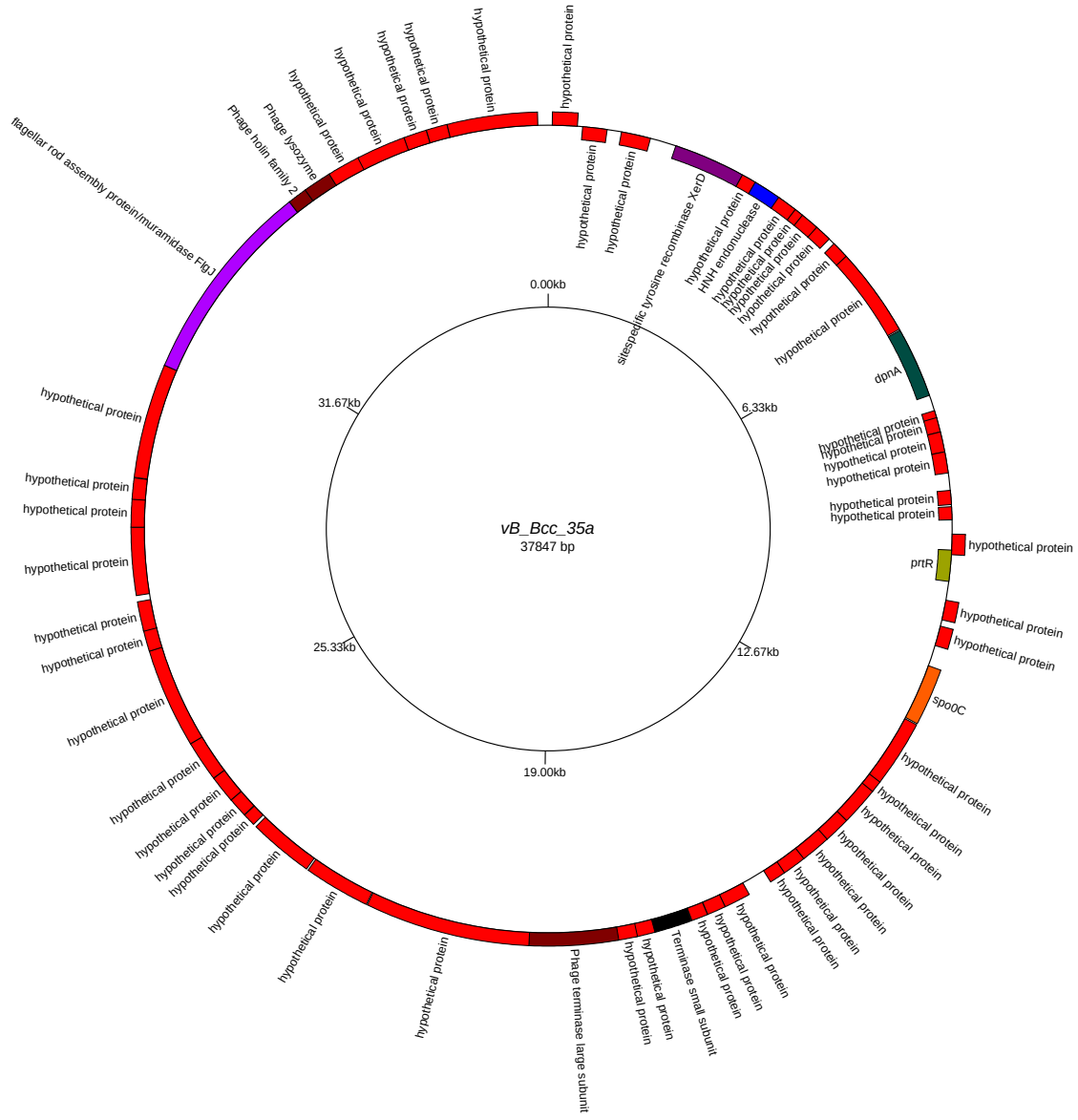


Figure 21: Genome map of vB_Bcc_32a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_35a genome size was 38.7 kb, the phage was assembled using SPAdes and has a mean coverage of 35. The genome map of this phage is shown in Figure 22. The total number of genes identified was 60 of which 10 were identified with putative functions these can be seen in Table 22. Blastn on viruses only database showed a 699 bp / 756 bp match with *Burkholderia* phage Bcep176 (GenBank: NC_007497.1). The phage vB_Bcc_35a had a 62.8 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
3021	1936	Reverse	site-specific tyrosine recombinase XerD
3661	3242	Reverse	HNH endonuclease
7429	6362	Reverse	Modification methylase DpnIIB
10247	9777	Reverse	HTH-type transcriptional regulator PrtR
11542	12387	Forward	Chromosome-partitioning protein Spo0J
16813	17373	Forward	Terminase small subunit
17899	19206	Forward	Phage terminase large subunit
30815	33817	Forward	flagellar rod assembly protein/muramidase FlgJ
33817	34104	Forward	Phage holin family 2
34101	34520	Forward	Phage lysozyme

Table 22: Functional genes identified for vB_Bcc_35a.


vB_Bcc_35a
37847 bp
0.00kb
6.33kb
12.67kb
19.00kb
25.33kb
31.67kb
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
Phage holin family 2
Phage lysozyme
flagellar rod assembly protein/muramidase FlgJ
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein

The temperate phage vB_Bcc_36a genome size was 45 kb, the phage was assembled using SPAdes and has a mean coverage of 1027. The genome map of this phage is shown in Figure 23. The total number of genes identified was 73 of which 11 were identified with putative functions these can be seen in Table 23. Blastn on viruses only database showed a 127 bp / 154 bp match with *Burkholderia* phage Bcep176 (GenBank: NC_007497.1). The phage vB_Bcc_36a had a 63.3 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
1582	3129	Forward	Terminase-like family protein
4648	5289	Forward	Phage Mu protein F like protein
13420	15207	Forward	Peptidoglycan hydrolase FlgJ
22908	24464	Forward	Pectate lyase superfamily protein
25311	25877	Forward	Chitinase class I
27266	27862	Forward	Lipase (class 3)
29149	28160	Reverse	site-specific tyrosine recombinase XerC
31841	31599	Reverse	Helix-turn-helix domain protein
35463	34783	Reverse	YqaJ-like viral recombinase domain protein
40753	40370	Reverse	Fic/DOC family protein
44437	44907	Forward	Endodeoxyribonuclease RusA

Table 23: Functional genes identified for vB_Bcc_36a.

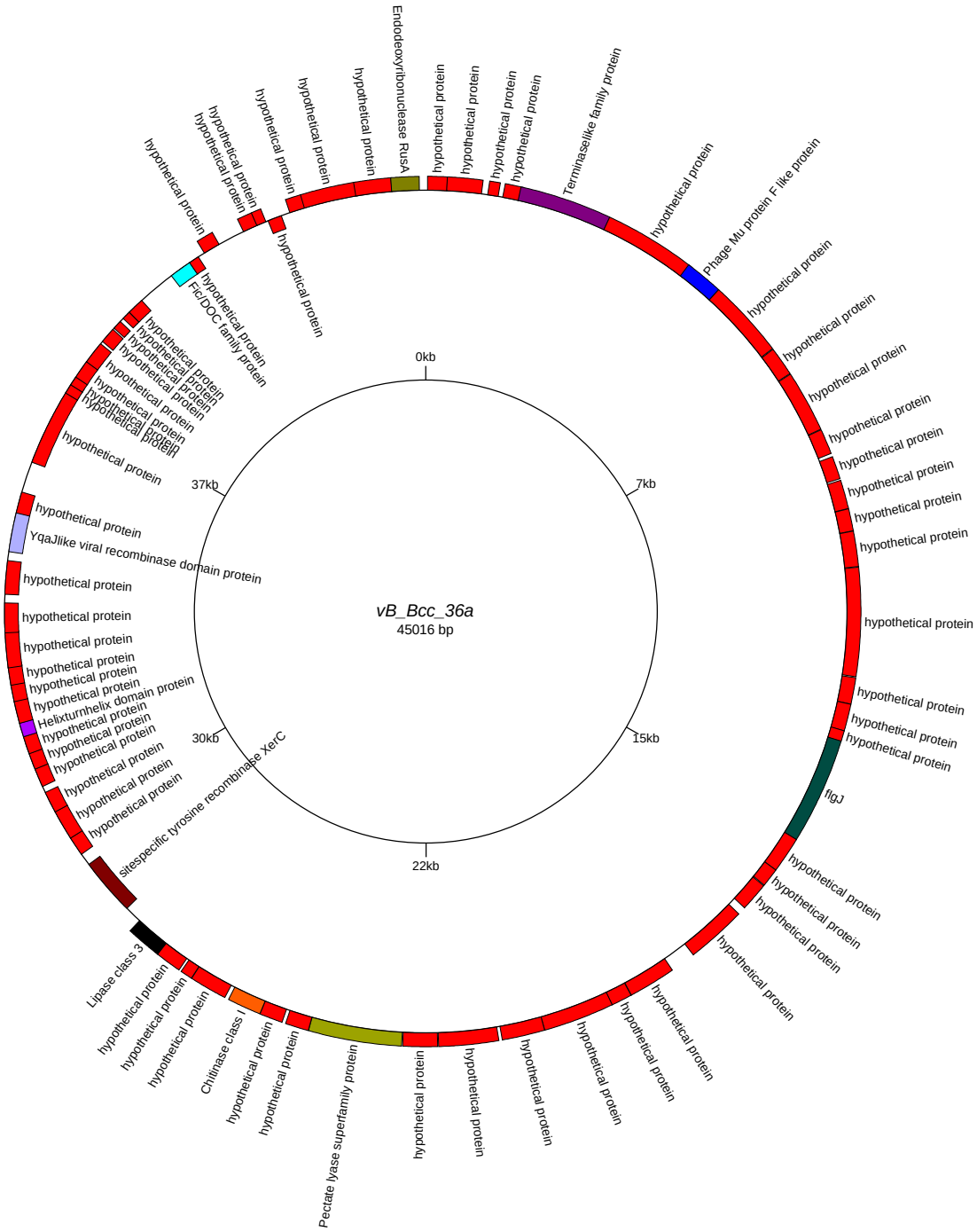


Figure 23: Genome map of vB_Bcc_36a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_39a genome size was 42.7 kb, the phage was assembled using SPAdes and has a mean coverage of 180. The genome map of this phage is shown in Figure 24. The total number of genes identified was 57 of which 20 were identified with putative functions these can be seen in Table 24. Blastn on viruses only database showed an 11.7 kb / 13.2 kb match with *Burkholderia* phage KS9 (GenBank: NC_013055.1). The phage vB_Bcc_39a had a 61.7 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
660	2324	Forward	Phage Terminase
2321	3610	Forward	Phage portal protein
3576	4379	Forward	ATP-dependent Clp protease proteolytic subunit 1
4448	5713	Forward	Phage capsid family protein
6359	6685	Forward	Phage head-tail joining protein
7994	8461	Forward	Phage tail assembly chaperone
8750	12850	Forward	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
12850	13188	Forward	Phage minor tail protein
14696	15379	Forward	Phage minor tail protein L
15429	16181	Forward	NlpC/P60 family protein
16178	16741	Forward	Bacteriophage lambda tail assembly protein I
21822	22106	Forward	Phage holin family 2
22109	22558	Forward	Peptidase M15
27365	26577	Reverse	Region found in RelA / SpoT proteins
28196	29209	Reverse	site-specific tyrosine recombinase XerD
30656	30348	Reverse	Helix-turn-helix domain protein
32541	31786	Reverse	thiamine biosynthesis protein ThiF
38300	37836	Reverse	Helix-turn-helix
39415	40782	Forward	Replicative DNA helicase
42339	42707	Forward	HNH endonuclease

Table 24: Functional genes identified for vB_Bcc_39a.

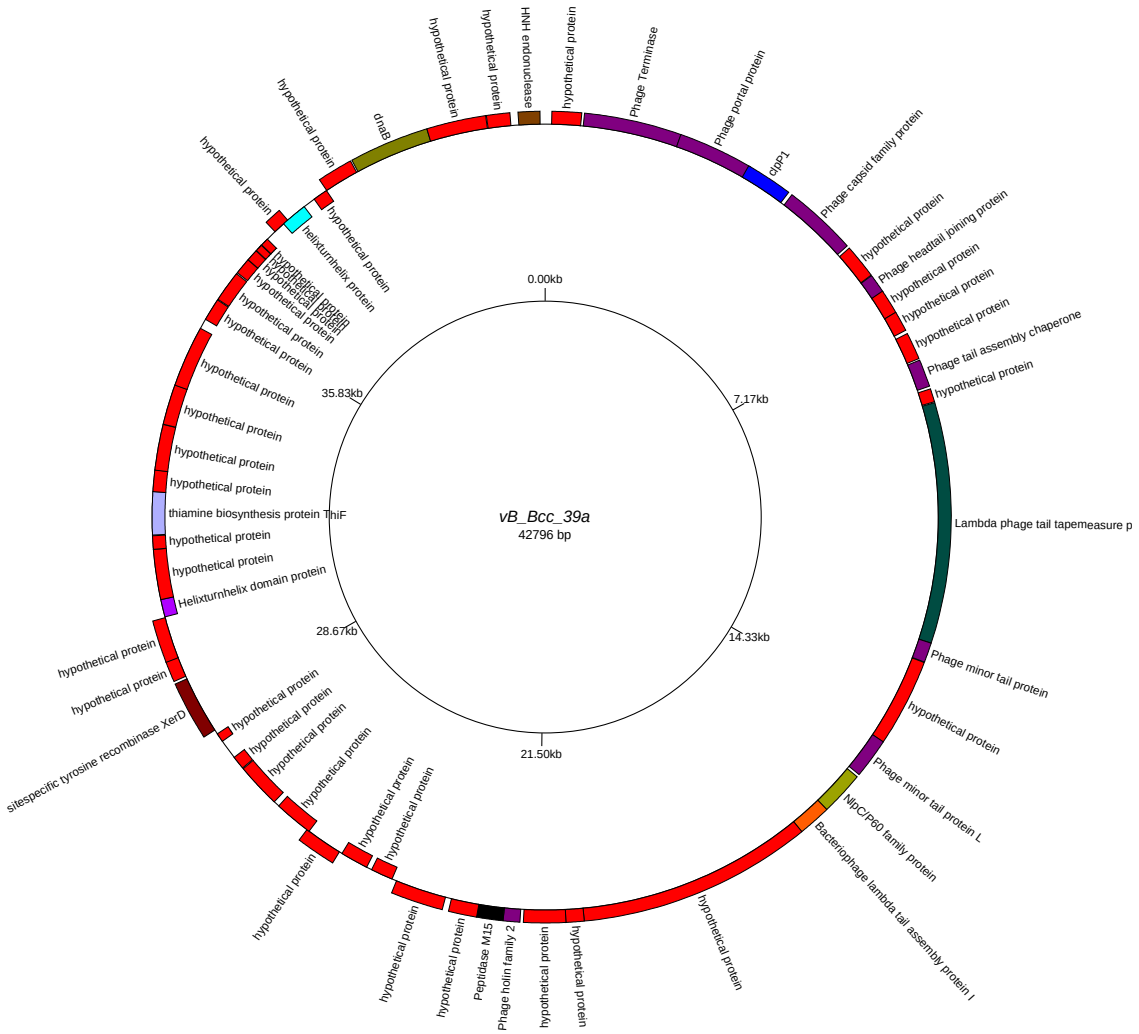


Figure 24: Genome map of vB_Bcc_39a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_43a genome size was 57.3 kb, the phage was assembled using SPAdes and has a mean coverage of 154. The genome map of this phage is shown in Figure 25. The total number of genes identified was 86 of which 16 were identified with putative functions these can be seen in Table 25. Blastn on viruses only database showed a 127 bp / 158 bp match with *Salicola phage* CGphi29 (GenBank: NC_020844.1). The phage vB_Bcc_43a had a 60.5 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
3199	4311	Forward	O-acetyltransferase OatA
5118	4300	Reverse	NTE family protein RssA
6807	6157	Reverse	Chitinase class I
12448	12044	Reverse	putative endopeptidase precursor
17373	14404	Reverse	Lambda phage tail tape-measure protein (Tape_meas_lam_C)
19197	18559	Reverse	Phage tail protein
20008	19643	Reverse	Phage head-tail joining protein
20636	20031	Reverse	Phage gp6-like head-tail connector protein
22623	21217	Reverse	Phage capsid family protein
23688	22780	Reverse	Putative signal peptide peptidase SppA
25163	23685	Reverse	Phage portal protein
27048	25363	Reverse	Phage Terminase
28044	27727	Reverse	HNH endonuclease
42559	41447	Reverse	Modification methylase DpnIIB
46507	47589	Forward	DNA polymerase III subunit beta
56596	55547	Reverse	Phage integrase family protein

Table 25: Functional genes identified for vB_Bcc_43a.

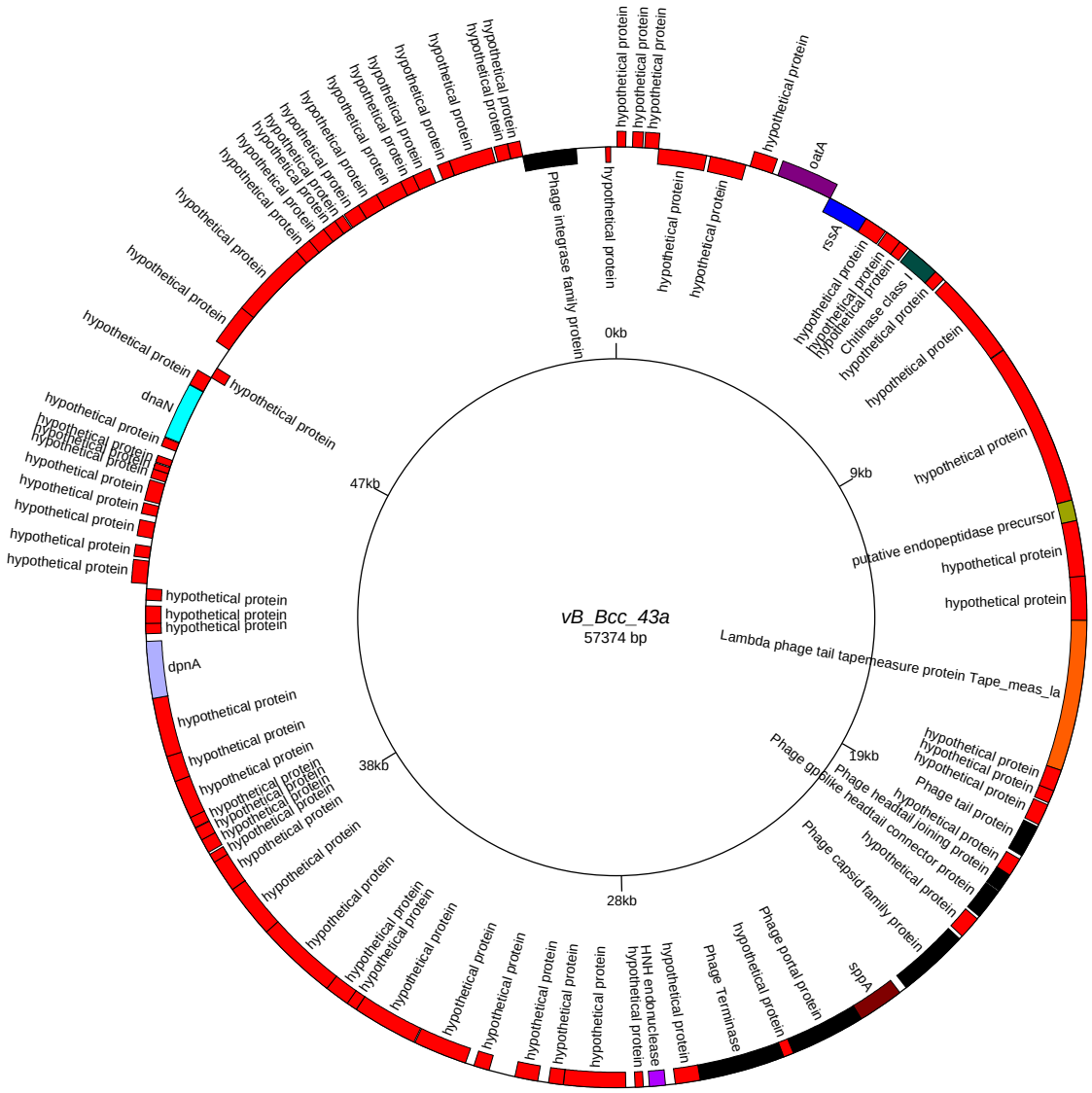


Figure 25: Genome map of vB_Bcc_43a produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.

The temperate phage vB_Bcc_43b genome size was 42.9 kb, the phage was assembled using SPAdes and has a mean coverage of 25. The genome map of this phage is shown in Figure 26. The total number of genes identified was 55 of which 22 were identified with putative functions these can be seen in Table 26. Blastn on viruses only database showed a 751 bp / 936 bp match with *Burkholderia* phage phiE125 (GenBank: NC_003309.1). The phage vB_Bcc_43b had a 62.1 % G-C content. No tRNA gene was identified however; several genes associated with structural and regulatory proteins were identified.

Start	End	Direction	Putative functional protein
633	2552	Forward	Phage terminase large subunit (GpA)
2769	4262	Forward	Phage portal protein, lambda family
4259	5326	Forward	ATP-dependent Clp protease proteolytic subunit
7062	7592	Forward	Prophage minor tail protein Z (GPZ)
8146	8790	Forward	Phage-related baseplate assembly protein
9050	9385	Forward	Gene 25-like lysozyme
9382	10278	Forward	Baseplate J-like protein
10268	10846	Forward	Phage tail protein (Tail_P2_I)
12868	13647	Forward	Caudovirales tail fibre assembly protein
13732	14901	Forward	Phage tail sheath protein
14912	15415	Forward	Phage tail tube protein FII
15896	18337	Forward	Phage-related minor tail protein
18348	19226	Forward	Phage P2 GpU
19201	19407	Forward	Phage Tail Protein X
19417	20466	Forward	Phage late control gene D protein (GPD)
20989	21555	Forward	putative Peptidoglycan domain protein
22477	23262	Forward	Modification methylase DpnIIA
29070	30227	Forward	Tyrosine recombinase XerC
30322	31461	Forward	ParB-like nuclease domain protein
33460	32507	Reverse	Modification methylase BspRI
38949	41456	Forward	DNA primase TraC

41677	42450	Forward	Chaperone protein DnaJ

Table 26: Functional genes identified for vB_Bcc_43b.

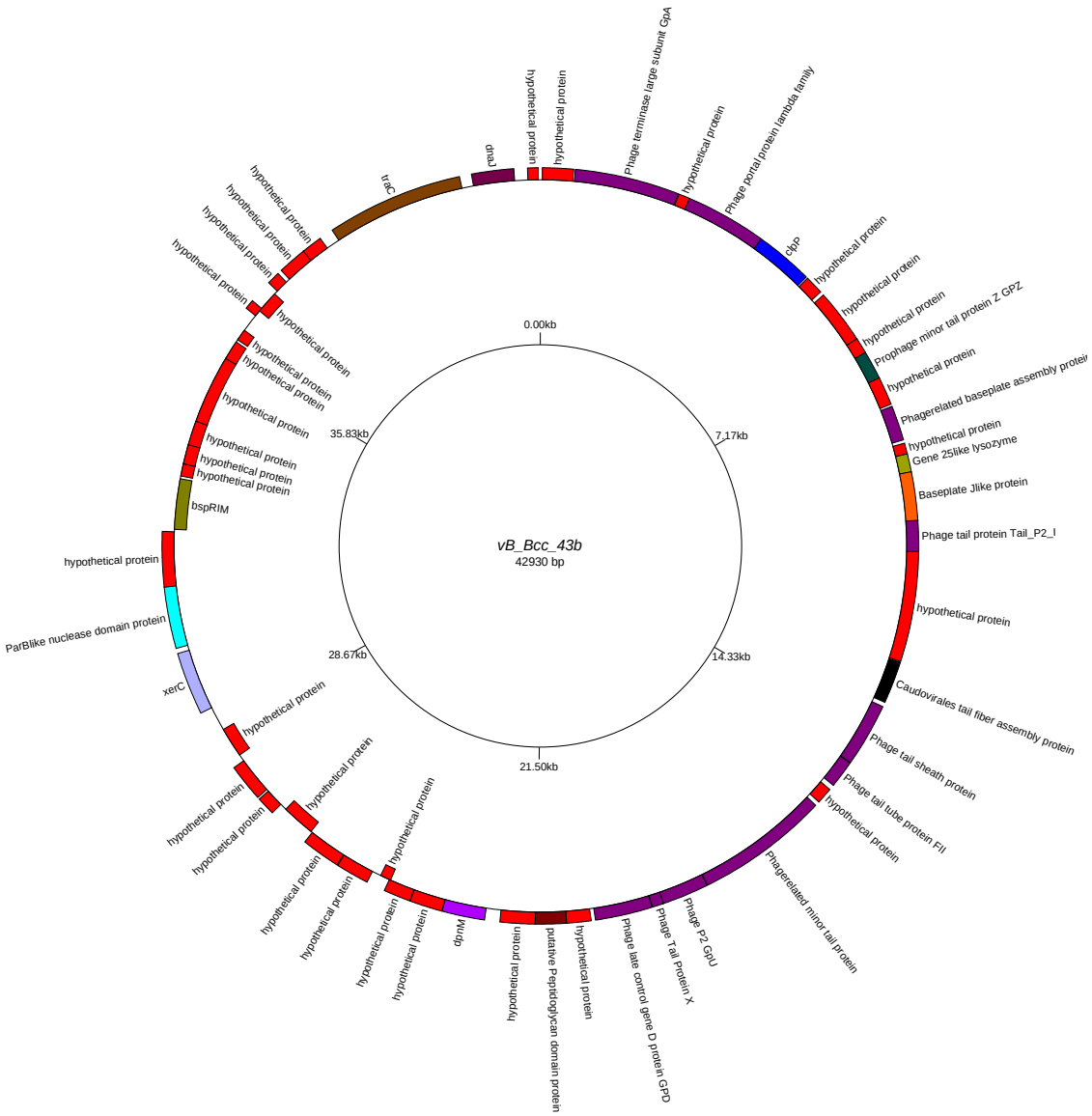


Figure 26: Genome map of vB_Bcc_43b produced using GenomeVx (http://wolfe.gen.tcd.ie/GenomeVx (Conant and Wolfe 2008)). Genes encoded on forward strand are shown outward and genes encoded on the reverse strand are shown inward.